Characterisation of a putative control element which lies between the imprinted IGF2 and H19 genes in the mouse

Charalambous, M.

Award date:
2000

Awarding institution:
University of Bath

CHARACTERISATION OF A PUTATIVE CONTROL ELEMENT WHICH LIES BETWEEN THE IMPRINTED *IGF2* AND *H19* GENES IN THE MOUSE

Submitted by M. Charalambous
for the degree of PhD
of the University of Bath
2000.

M. Charalambous

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ABSTRACT

The imprinted *Igf2* and *H19* genes lie within a 100kb region of mouse chromosome 7 and its syntenic region 11p15.5, disruptions of which have been implicated in the paediatric overgrowth disorder Beckwith Wiedemann syndrome. Both genes are expressed co-ordinately in a wide range of embryonic and extraembryonic tissues from a single parental allele. The *Igf2* gene is expressed predominantly from the paternally derived allele, and *H19* from the maternally derived allele. The mechanisms by which the two genes achieve their spatial, temporal, and allele-specific expression patterns are not fully understood, but are thought to be at least partially dependent upon *cis*-acting genomic elements which lie within the immediate chromosomal domain.

A candidate for such an element, the centrally conserved domain (CCD) was initially identified in a study that scanned approximately 130kb of the genomic region containing the *H19* and *Igf2* genes for regions that were sensitive to the nuclease DNaseI. A region 32kb upstream of the *H19* gene was found to display a high level of nuclease sensitivity on both parental alleles. This region was also found to be hypomethylated in comparison to surrounding DNA, and to be conserved in a number of mammalian species. This region could function as a tissue specific enhancer or silencer, or a further imprinting element at this locus. The purpose of this work is to investigate the role of the CCD in the tissue-specific expression and imprinting of the *H19* and *Igf2* genes, by utilising a panel of transgenic mice bearing putative and known *Igf2/H19* control elements fused *in cis* with the reporter gene firefly *luciferase*. *Luciferase* reporter assays performed on neonatal transgenic mice have revealed that this region can drive reporter gene expression in the brain, specifically in the exchange tissues. In the brain, *Igf2* gene expression is concentrated in the exchange tissues, i.e., the choroid plexus and leptomeninges. This expression is from both parental alleles, and persists into adult life.

A tissue culture system, primary culture of cells derived from the mouse choroid plexus, has been set up in order to dissect the function of the CCD *in vitro*, and preliminary work done to derive an immortal cell line from these cells. Such a cell line may form a model system by which to study how *Igf2* escapes imprinting in the choroid plexus and leptomeninges of the brain.

ACKNOWLEDGEMENTS

I would like to thank Dr Robert Costa (University of Illinois at Chicago) for the *pTTR-7KbexV3* plasmid containing *transthyretin* gene sequences. The samples of the muscle cell line *C2* were donated by Dr Janet Smith (University of Birmingham) who also provided the protocol for antigen retrieval protocol. Prof. M. Azim Surani, Dr Justin Ainscough (University of Cambridge), and Dr Tsuyoshi Koide (National Institute of Genetics, Shizuo, Japan) provided the 2kb CCD clone that is the focus of this work, and also generously donated sequence information and results prior to publication. Dr Alan Clarke (*CRC* laboratories, University Medical School, Edinburgh) was the source of the *p53* null mice. Dr Betsy Pownell and Dr Harv Isaacs (University of York) kindly taught me how to perform ribonuclease protection experiments. Dr David Tosh and Kenny Shen (University of Bath) performed the confocal microscopy, were a source of wisdom concerning cell culture techniques, as well as being good company during endless tissue culture. Prof. Laurence Hurst (University of Bath) performed the comparative sequence analysis of the mouse and rat CCD regions, which is reproduced in this report. Dr Ghislaine Dell created the *LacZ* reporter constructs, and deserves eternal gratitude for help with all those e14.5 dissections. Thanks to Dr Bill Bennett for assistance with all matters technical and culinary, and especially for performing a large proportion of the embryo transfers, and to Dr James Dutton for teaching me almost everything I know about molecular biology. Dr Andrew Ward was/is equally an excellent supervisor, an inspiring discussion partner, and a superb mouse technician. Finally, thanks to all members of lab. 0.76, past and present, for providing a relaxed atmosphere in which to work, most especially to Angela Pauliny and Derek Paisley, for welcome distractions. This work was funded by a studentship from the School of Biology and Biochemistry, University of Bath.

This work is dedicated to Penda,

"Never stay up on the barren heights of cleverness, but come down into the green valleys of silliness."
Ludwig Wittgenstein, 1948

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Air	*Antisense insulin-like growth factor 2 receptor transcript*
AS	Angelman Syndrome
bp	base pair
BSA	bovine serum albumin
BWS	Beckwith Weidemann syndrome
CCD	centrally conserved domain
D(X)	X days post-partum
DMR	differentially methylated region
DNA	deoxyribonucleic acid
Dnmt1	*De-novo methyltransferase 1*
e(X)	embryonic day (X)
ECM	extra-cellular matrix
ES cells	embryonic stem cells
FCS	foetal calf serum
H(X)	histone (X)
HDAC	histone deacetylase
HS	hypersensitive site
IC	imprinting centre
Igf2	*Insulin-like growth factor 2*
Igf2r	*Insulin-like growth factor 2 receptor*
kb	kilobase
LCR	locus control region
LINE	long interspersed repeat element
MatDi7	maternal disomy of chromosome 7
Mb	megabase
MeCP2	5' methyl-cytosine binding protein 2
MNase	micrococcal nuclease
mRNA	messenger RNA
NTG	non-transgenic
ORF	open reading frame

P0-P4	promoter 0-4
PAC	P1 artificial chromosome
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PGC	primordial germ cell
PMEF	primary embryonic fibroblast
Pol II	RNA polymerase II
PWS	Prader-Willi Syndrome
RNA	ribonucleic acid
SCP	sheep choroid plexus
TRD	transcriptional regulation domain
TSA	trichostatin A
TTR	*transthyretin*
Ubx	*ultrabithorax*
UPD	uniparental disomy
YAC	yeast artificial chromosome

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Imprinting

General

Nuclear transplantation experiments that produced mouse embryos with only maternal or paternal chromosomes, while maintaining the normal diploid number, revealed that such embryos are not viable [1, 2, 3]. These experiments showed that both parental genomes are required for normal development. The cause of this 'non-complementation' phenomenon has been found to be due to the existence of imprinted genes, i.e., those genes whose expression status depends upon their parental origin. To date, more than 30 imprinted genes have been discovered in the mammalian genome, with a large proportion of these genes conserving their imprinted status between mouse and man [4]. It is unclear whether all imprinted genes share a common mechanism by which they acquire a parental allele-specific mode of expression, or whether different genes within this class have manipulated different epigenetic systems within the cell to reach the same end. Each imprinted gene identified to date shares one or more of a set of features which has come to characterise imprinted genes. These features, and their implications in the understanding of the mechanism/s of imprinting are discussed below.

Features of imprinted genes

Imprinted genes as a class have been suggested to share several common features, including; association with regions of parental-allele specific methylation and/or chromatin structure, the presence of antisense or overlapping transcripts, linkage to G-rich repeat sequences, the tendency to be clustered in chromosomal domains, as well as in regions with allele-specific differences in DNA replication timing, and the association of such genes with a role in embryonic growth.

DNA methylation

The necessity of correct genome methylation from early development for the correct allele-specific expression has been demonstrated for most imprinted genes. One important exception is the *Mash2* gene, where the manipulation of levels of the major *de-novo* methyltransferase (*Dnmt1*) gene has no effect upon its imprinted expression [5]. Many

imprinted genes contain, or are flanked by regions that are differentially methylated in an allele-specific manner (e.g., the insulin-like growth factor 2 (*Igf2*) gene [6, 7, 8], *H19* [9], the insulin-like growth factor 2 receptor (*Igf2r*) gene [10], *KvLQT1* [11], *SNRPN* [12]). The role of methylation in the imprinting mechanism is discussed in greater detail below.

Antisense transcripts

Antisense transcripts overlapping the coding sequence of several imprinted genes have been identified at several loci, including *Igf2r* [13], *Igf2*, [6, 14], *KvLQT1*, [15, 11], *UBE3A* [16], *Xist* [17] and reviewed in [18]. In all cases these antisense RNAs are imprinted themselves, and with the exception of the *Igf2* antisense transcript, in the opposite direction to the coding genes which they overlap. As these antisense RNAs also tend to be expressed in the same tissues as their sense equivalents, it has been suggested that they may exist to regulate these genes by a competition mechanism [19], i.e., that a cell can only transcribe the sense or the antisense gene at a particular locus, and that the imprint acts to dictate which of the two transcripts will 'win', in an allele specific manner. Such a mechanism can not explain how the parental alleles are distinguished, and it has not formally been demonstrated that transcription of a gene will downregulate expression from an overlapping partner. In fact, the *Igf2r* antisense transcript (*Air*) overlaps but does not imprint the neighbouring *Mas1* gene [20].

Clustering of imprinted genes

The tendency of imprinted genes to be clustered raises the question of whether a common 'imprinting control region' is required to regulate the allele-specific regulation of multiple genes *in-cis*. Such an imprinting centre appears to exist to regulate the genes at the Prader-Willi/Angelman Syndrome (PWS/AS) region on human chromosome 15 ([21, 22], and references therein). 15q11-q13 contains at least four paternally expressed genes, *SNRPN, ZNF127, NDN* and *IPW* as well as two paternally expressed sequences, *PAR1* and *PAR5*. Absence of paternally inherited 15q11-q13 causes PWS. This region also contains one maternally expressed gene, *UBE3A*, mutations in which cause AS. Co-ordinate control of imprinting at this cluster is caused by the imprinting centre (IC), which is also functionally conserved in the mouse. This centre has been localised to a 100 kilobase (kb) region containing the *SNRPN* gene, which is deleted in both PWS and AS patients. This

element may have a bipartite structure, as in AS families the smallest deletion overlap is approximately 1.2kb and maps to 40kb upstream of *SNRPN,* whereas in PWS families the common deletion spans the *SNRPN* gene [18]. The IC controls a large imprinting domain, and mutations in this region appear to block the resetting of the imprint in the germline [23], as well as the maintenance of the paternal epigenotype [24].

Another intensively studied imprinting cluster is the Beckwith-Weidemann Syndrome (BWS) region on human chromosome 11p15.5, and its syntenic region in the mouse on distal chromosome 7. This cluster in the mouse spans approximately 1Mb, and contains at least 10 imprinted genes, four of which are paternally expressed (*Igf2, Igf2AS/PEG8, Ins2, LIT1),* and the remainder of which are maternally expressed (*H19, Mash2, p57^{KIP2}, Ipl* and *KvLQT1*, [25], and see references therein). The human cluster is similar in size, and the relative transcriptional orientation of the genes appears to be conserved between the two species [25]. The cluster contains within it several genes which have been found not to be imprinted in all tissues analysed (*TSSC4, TSSC6,* [26]), though the tissue, and developmental specificity of the imprinting status of most imprinted genes makes it problematic to rule out that these genes may show allele specific expression in some tissues. The BWS cluster does not appear to contain a single imprinting centre, though deletion of a 13kb region including the *H19* gene can lead to loss of imprinting of the linked *Igf2, Ins2* genes [27], see below), but not further genes in this cluster [5]. Mutations in the *KvLQT1* and the overlapping *LIT1* gene have been implicated in the loss of imprinting of *IGF2* in some cases of BWS [11], and this region has also been proposed to direct the maternal expression of the neighbouring *p57^{Kip2}* gene [28]. While a single IC has not been characterised in this region, neighbouring genes are clearly co-ordinately regulated to some extent.

Analogous to the PWS/AS imprinting centre is the X-inactivation centre (Xic). Deletions in this region of the X chromosome prevent X inactivation *in cis,* leading to inappropriate expression of genes along the chromosome (reviewed in [29]). Silencing of the inactive X chromosome is reliant upon the spread of constitutive heterochromatin from the inactivation centre, and is dependent upon transcription of the *Xist* gene which maps within this region [30]. The *Xist* transcript coats the inactive chromosome [31], though *Xist* RNA levels are relatively low (approximately 100 molecules/cell) in some tissues, requiring a

single *Xist* molecule to inactivate approximately 1% of the X chromosome [29]. Binding of *Xist* is not thought to be sufficient to silence genes over such distances, therefore Brockdorff et. al. [29] have proposed that the spread of heterochromatin is aided by 'booster' elements or 'way stations', i.e., sequences that facilitate the spread of higher-order chromatin packaging. Lyon [32] has proposed that long interspersed repeat elements (LINE)-1 (L1) as a candidate for these 'booster' elements. Bailey et. al. [33] have studied the distribution of L1 repetitive elements on human chromosomes, and found that the X chromosome is enriched 2-fold for these elements (as compared to autosomes). Furthermore, regional analysis of the X chromosome revealed that the most significant clustering of L1 sequences is at the X inactivation centre, and regions harbouring genes that escape X inactivation are significantly poorer in these repetitive elements.

Repetitive sequences

Interestingly, imprinted genes are often associated with repetitive DNA (*Igf2* and *Ins2* [6], *H19* [34], *Igf2r* [20], *KvLQT1* [15], $p57^{Kip2}$ [35]), suggesting that such regions are required in imprinted domains, perhaps to stabilise gene silencing. In addition, there appears to be a necessity for multiple elements to stabilise the imprinting mechanism. Transgenes constructed containing elements from imprinted gene regions generally only reproducibly exhibit stable patterns of imprinted gene expression either when they are very large [36, 13]), or present at high copy number [37, 38, 39]. The presence of 'booster' elements, which have a partial but cumulative effect upon the maintenance of an epigenetic state, is reminiscent of the mechanism of *Polycomb*-mediated gene silencing in *Drosophila*. At the *Ultrabithorax* (*Ubx*) gene, for example, binding sites for *Polycomb*-group proteins are dispersed over a wide area, and act in combination to create an epigenetically silent domain (in the appropriate tissues), encompassing their target gene [40]. It is interesting to note that the imprinting status of many genes is also tissue specific, and that *cis* factors that mediate gene silencing in such contexts must therefore be tissue responsive.

Replication timing

A further feature of imprinted genes is their tendency to manifest allele-specific differences in DNA replication timing at S-phase. At the PWS/AS cluster, at the *Igf2* locus

and the *Igf2r* locus, in both mouse and humans, the paternal chromosome replicates earlier in S-phase than the maternal chromosome [41, 42]. Asynchronous replication of these chromosomal domains is evident from the earliest stages of embryogenesis, and persists throughout development [43]. The establishment of allele specific replication timing occurs in the gametes, preceding fertilisation, therefore the events that mark the cell cycle specific-timing of DNA replication in imprinted regions must reflect an early event in the formation of an imprinting mark. However, this mark is not allele-specific, as both paternally and maternally imprinted genes reside in regions that are early replicating on the paternal chromosome. Asynchronously replicating regions may mark areas of the genome that are competent for monoallelic gene expression, as the randomly inactivated X chromosome is late replicating (reviewed in [44]), as are inactive alleles of the monoallelic (but non-imprinted) olfactory receptor loci [43]. *Cis* deletions that disrupt the imprinting of neighbouring genes (such as PWS/AS IC [42]), and the *H19* gene region [45]) lead to synchronous replication, providing a direct link between imprinting and asynchronous replication timing. It is not known whether this phenomenon is a necessary feature of the mechanism of monoallelic expression i.e., it has been suggested that DNA binding factors necessary for monoallelic expression may only be present in a restricted time-window of the cell cycle [46]. Alternatively, asynchronous replication of parental alleles could be a secondary effect of changes in chromatin structure that accompany imprinted genes. Manipulation of chromatin structure, brought about by treatment of cells with histone deacetylase inhibitors, leads to synchronous replication of imprinted gene regions [47].

Functional relatedness and imprinting theories

The products of many imprinted genes play a role in growth, either by effects on cell proliferation (*p57^{Kip2}* [35], *NOEY2* [48], *ZAC/PLAGL1* [49]); embryonic and placental growth (PWS/AS genes [21], *Igf2* [50], *Igf2r* [51], *Mash2* [52], *Peg/Mest* [53, 54, 55], *Esx1* [56]; and/or a role in the insulin growth factor signalling pathway (*Igf2* [50], *Igf2r* [51], *Grb10* [57], *Ins1, Ins2* [58] and *ZAC/PLAGL1* [49]. The superficial similarity of the type of genes that tend to be imprinted has led many to assume that there must be a connection between function of imprinting and growth.

An organism that preferentially inactivates a subset of its genes becomes functionally haploid at those loci. Spencer et. al. [59] have shown by mathematical modelling that without some direct advantage to imprinting, any modifier gene causing its target locus to be imprinted, but having no direct fitness benefits will be eliminated from a population. A number of theories have been proposed to describe the evolution of imprinting, many of which can be rejected on the basis that they do not explain the available facts (such as growth effects, parental direction of imprinting, species specificity, etc), (reviewed in [60]).

Early observations of uniparental disomies (UPD) in mice and humans (such as human paternal UPD 11p15.5) suggested that paternal UPD tended to lead to increased embryonic growth, whereas individuals with maternal UPD tended to be smaller in size throughout development. This led Moore and Haig [61] to propose one of the most widely accepted theories of the evolution of imprinting, the parental conflict model. This model states that, in a species where a female takes more than one mate, there is a conflict between the strategies employed by males and females to maximise their reproductive fitness. As the female is equally related to all of her offspring, she will provide each of her brood sufficient resources to survive, without compromising her future reproductive potential. The father, however, has no stake in his current mates' future offspring, so will try to divert maternal resources to his offspring, even at a large cost to his mate. Moore and Haig [61] stated that if a locus has preferential paternal expression, it will function to increase nutrient demands on the mother, whereas maternally expressed genes will act to decrease these demands. The success of this model lies in its ability to predict the existence of imprinting over random gene inactivation, the role of imprinted genes in growth pathways, and the species specificity of this phenomenon.

The major prediction of the conflict model is that paternally expressed genes should promote growth, whereas maternally expressed genes should suppress it. This prediction was certainly borne out for the first imprinted genes discovered (*Igf2* [50], and *Igf2r* [51]), and the early uniparental disomy data (reviewed in [61]). A comprehensive study of the growth effects associated with uniparental disomies [62], concluded that much of the data did not support the conflict hypothesis. While maternal UPDs did tend to be growth suppressing (as would be expected from the absence of paternal growth enhancers, or a double dose of

maternal growth suppressors), so also did paternal UPDs. However, the authors did suggest that, as imprinted genes occur in clusters, the dramatic changes in gene expression brought about by alterations in the imprinted status of several genes might well have complicated the analysis. Studies of the effects of deletions in individual imprinted loci also do not always produce the outcome predicted by the parental conflict model. While some paternally expressed genes are growth enhancing (e.g. *Igf2* [50], *Mest* [54] (though the adult behavioural phenotype is unexpected), and *Peg3* [63]); in the case of some maternal genes, although implicated in cell proliferation (e.g., $p57^{Kip2}$ [35], *NOEY2* [48]), their loss of function has no overt growth phenotype.

Moore and Haig [61] have suggested that one of the main sites for parental conflict should be the placenta, as this is the site for acquisition of maternal resources by the foetus. For the parental conflict model to be supported, imprinted genes expressed in the placenta should act to divert resources to the embryo. The imprinted expression of two placental genes appear to confound this assumption, as maternal deletions of *Mash2* [52] and the X-linked *Esx1* [56] lead to placental malformation and growth retardation of the embryo. The implication that the expression patterns of these genes contradicts the conflict model could be due to a poor understanding of the relationship between placental and foetal growth. The notion that large placentas are associated with large embryos has been shown not to hold for both some cases of paternal UPD, (which have large placentas but small embryos, [60]), and in experiments that manipulate placental size by generating interspecific crosses of mouse strains. Some intercrosses between *Mus musculus* and *Mus spretus* show a large variation in placental size [64], which is not correlated to embryo size. The authors conclude, however, that the severe morphological changes present in many large placentas may skew the analysis.

The conflict model predicts that imprinting should be confined to those species where offspring are nourished directly from maternal tissues. This prediction is largely supported by observation, imprinting is found in mammals and angiosperms, as well as marsupials, and parthenogenesis has been reported in all major groups of vertebrates except mammals [61]. However, parental allele-specific differences in gene expression have been reported in *Drosophila* [65], [66], and in the zebrafish *Danio rerio* [67], so while both

parental genomes are not required for viability in these species, imprinting mechanisms may exist.

Further criticisms of the conflict model include the observation that imprinted genes are not fast evolving, as might be expected if an 'arms race' existed between maternal and paternal genomes [68]. Furthermore, the theory fails to explain why some, but not all embryonic growth factors are imprinted [60]. In spite of this, the conflict model provides a close fit with many of the observed features of imprinted genes.

The Igf2/*H19 gene pair*

Igf2 and *H19* lie within a 100kb region at the telomeric end of mouse chromosome 7, and in humans in the syntenic region 11p15.5, within a large cluster of imprinted genes ([25], and references therein). The two genes are also closely linked in the marsupial (the tamar wallaby, *Macropus eugeni* [69]).

Igf2

IGF2 is a mitogenic growth factor, structurally related to the hormone insulin. The *Igf2* gene is expressed from early development in a wide range of embryonic and extraembryonic tissues (see below), in most cases only from the paternally-derived chromosome [70]. Loss of expression of *Igf2* leads to a marked reduction in growth. Mice carrying a paternal deletion of this gene are proportionate dwarfs, of approximately 60% of the body weight of their wild-type littermates [50]. Overexpression of *IGF2* in humans, due to inappropriate activation of the maternal allele, has been associated with the fetal overgrowth syndrome Beckwith-Weidemann Syndrome. Mouse models in which *Igf2* is overexpressed display some of the diagnostic features of these syndromes, such as pre- and postnatal overgrowth, polyhydramnios (excess fluid in the amniotic cavity), disproportionate organ overgrowth including macroglossia, and skeletal abnormalities [71]. The full range of phenotypes for this syndrome can be demonstrated in a mouse model in which *Igf2* is overexpressed, and the linked $p57^{Kip2}$ tumour suppressor gene function is lost [72, 73]. In addition, loss of imprinting, and biallelic expression of *IGF2* has been found in over 70% of solid tumours where examined (reviewed in [74]).

The mouse and human *Igf2* genes have a similar gene structure, with expression from four promoters (named P1 to P4 in humans, P0 to P3 in mouse, see ***Figure 1***) [75]. The

human P1 and exon 1 have no mouse homologue, and transcripts initiating from this site are found in adult liver [76]. In the mouse, no postnatal liver expression has been detected. The P0 promoter in the mouse is the site of initiation of a placental-specific transcript [6]. The remaining promoters are highly conserved between mouse and humans, and transcripts arising from them show a similar pattern of spatial and temporal expression during development. The coding sequences for the *IGF2* peptide are contained in the three 3' exons (see ***Figure 1***), and the structural gene is highly conserved across the two species [75].

H19

The *H19* gene encodes one of the most abundant RNA polymerase II transcripts in the mouse embryo. *H19* transcripts accumulate to greater than 1% of total mRNA in some foetal tissues [77].

The mouse *H19* gene consists of five exons and four very small introns (see ***Figure 1***), and encodes a 2.5kb transcript that is spliced and polyadenylated [78]. The human gene is very similar in structure to the mouse gene [77]. There is some evidence, however, that an allele-specific splice variant exists in humans, which results in the excision of exon 4 [79].

H19 lies *in cis* to *Igf2* on mouse chromosome 7 [80]. *H19* is oppositely imprinted to *Igf2*, i.e., it is transcribed from the maternally derived allele only, in the majority of tissues in which it is expressed [81].

The *H19* gene product is generally thought to function as an RNA. Although several open reading frames can be found within the mouse sequence, none of these are conserved in the human sequence. Despite this, a generally high level of sequence conservation can be found between the two species, particularly across the 3' region of exon 1, and in exon 2, which is also highly conserved in rat [82], and chicken [77]. Additional evidence that *H19* is not translated comes from a report that states that though *H19* mRNA is present in the cytoplasm, it does not associate with ribosomes. Instead the mRNA is though to co-localise with a high molecular weight protein fraction of the cytoplasm [77]. A conflicting study [83] asserts that *H19* mRNA does associate with polysomes in a variety of cell types in both mouse and human, but suggests that here the function of the RNA may be to regulate *IGF2* mRNA, with which *H19* mRNA co-localises.

Figure 1. Schematic diagram of the ~100kb mouse genomic domain containing the *Igf2* and *H19* genes. Grey boxes represent exons of the genes, *Igf2* contains eight exons in mouse, including upstream exons 1 and 2 (u1 and u2), and six additional exons. The *H19* gene contains five exons separated by four small introns. The four *Igf2* promoters (P0-P3), and single *H19* promoter are shown as horizontal arrows. Vertical arrows above the line represent regions of nuclease hypersensitivity, and horizontal black bars below the line represent regions of differential CpG methylation (DMRs). A detailed description of these regions, and the relevant references are contained within the text.

The regions within this domain which are present in *luciferase* reporter gene constructs analysed in this work, are highlighted below, and discussed within the text. In order these elements are: DMR (*Igf2* DMR1, approximately 2.8kb *EcoRI-BamHI* fragment); P3 (*Igf2 promoter 3,* nucleotides -162 to +74 with respect to the trancription start); CCD (a nuclease hypersensitive region lying midway between *Igf2* and *H19,* 2kb *EcoRI-EcoRI* fragment); the *H19* promoter, (a 2kb *BamHI-NheI* fragment); the *H19* enhancers (2.7kb SpeI-BglII fragment).

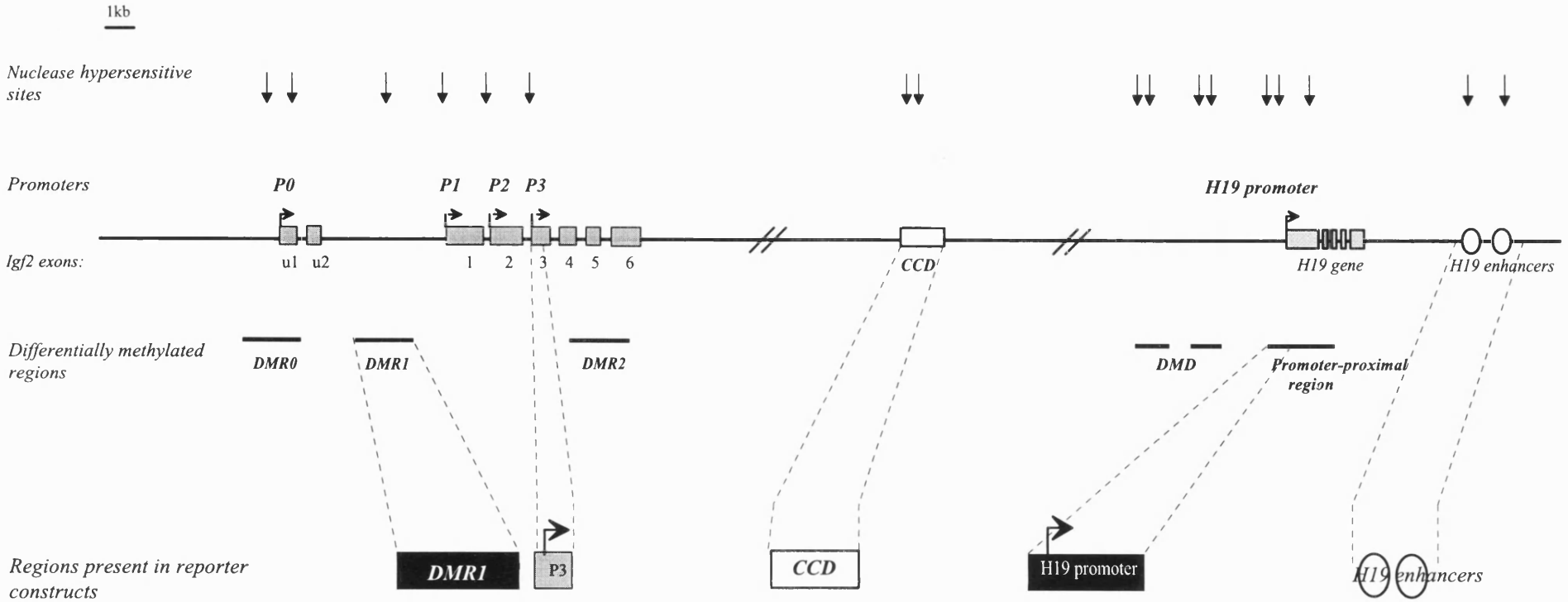


Figure 1. The mouse *Igf2-H19* genomic region.

At present the role of the *H19* gene in development remains elusive, but its abundance and strict sequence conservation [82] argue that this gene must be functional. Despite this, loss of expression of the *H19* coding region [84, 85], or overexpression of the gene from multicopy transgenes [37, 38, 39], have no obvious phenotype.

Tissue specific expression

The *Igf2* and *H19* genes are expressed broadly and at high levels in the developing embryo, and in the extraembryonic tissues. While strikingly similar in their tissue specificity, differences exist in the expression patterns of these two genes. Both the similarity and the differences in the expression patterns of the *Igf2* and *H19* genes during embryogenesis are discussed below.

Extraembryonic gene expression

Basal levels of *Igf2* RNA can be detected from the earliest stages of pre-implantation mouse development [86]. Abundant expression of *Igf2* is first detected in the trophectoderm at embryonic day 5.5 (e5.5). The first expression is seen in the polar trophectoderm (extraembryonic ectoderm and ectoplacental cone), with the mural trophectoderm lagging slightly behind. The first derivative of the epiblast to express *Igf2* is the extraembryonic mesoderm (e6.5-e7.0), which will form the allantois and visceral yolk sac. In these tissues, *Igf2* gene expression is detected from the earliest stage of formation of the amniotic folds [87]. At e9.5 the rodent yolk sac placenta converts into the chorioallantoic placenta. During this process the allantois fuses with the roof of the ectoplacental cone to form the allantoic plate, and later the labyrinthe. From this stage the allantoic mesoderm and constituent blood vessels emerge as the strongest expressers of *Igf2* in the murine placenta [88]. By e12.5 the labyrinthine trophoblast expresses *Igf2* at greater levels than the spongiotrophoblast, where expression declines completely as a subset of these cell differentiate into glycogen cells and secondary giant cells. From e14.5, the glycogen cells progressively infiltrate the maternal decidua, becoming concentrated around the large maternal arteries. These cells retain *Igf2* expression until birth. This extraembryonic expression pattern of *Igf2* is also observed in the rat, where the predominant sites of expression of *IGFII* are in the allantoic plate, and later the labyrinthine zone; and in the glycogen and giant cells that form from the spongiotrophoblast [89].

In human development high level induction of *IGF-II* mRNA again coincides with implantation ([90], [91]), where the gene is expressed in the trophectoderm. *IGF-II* mRNA continues to be expressed in the proliferative regions of the placenta, i.e. in the cytotrophoblast shell and intermediate trophoblasts (which are functionally equivalent to the murine early spongiotrophoblast and glycogen cells, respectively [88]) throughout development [92].

In contrast to *Igf2*, *H19* gene expression cannot be detected until late blastocyst stages. At e4.5 low levels of mRNA are transcribed in the trophectoderm. Following implantation, *H19* expression increases sharply, in the ectoplacental cone, the trophoblast giant cells, the extraembryonic ectoderm and the extraembryonic endoderm. *H19* mRNA cannot be detected in the allantoic folds of the extraembryonic mesoderm, which express high levels of *Igf2* mRNA. In subsequent stages of development, however, *H19* gene expression is very similar, if not identical, to that of *Igf2* [93]. The expression of *H19* in human placental development also closely resembles that of *IGFII* [94].

Embryonic gene expression

The embryonic expression patterns of *Igf2* and *H19* have mostly been studied in the mouse and the rat, though the expression patterns of these genes in the human foetus are broadly similar [94].

The first *Igf2* transcripts detected in the embryo proper in mouse are found at late primitive streak stage (e7.5) in the developing anterior and lateral mesoderm, which is continuous with the extraembryonic mesoderm. By e8.0, the strongest *Igf2* expression is found in the developing heart, and the head mesenchyme. The first endodermal expression in the embryo is observed at this stage, with significant expression of *Igf2* mRNA in the lining of the foregut [87].

The late-embryonic expression pattern of *Igf2* is most comprehensively documented in the rat [95], though this expression is at least superficially equivalent to the mouse [70]. In mesoderm, *Igf2* transcripts can be detected in all three subdivisions of the somites (sclerotome, dermatome, and myotome) from e10. As development proceeds in myotome, *Igf2* transcript levels are high throughout the formation of myoblasts and their differentiation into myotubules. In sclerotome, *Igf2* mRNA levels increase as mesodermal cells differentiate into cartilage, and remain high in chondrocytes, but transcript levels decrease dramatically immediately prior to ossification. Further mesodermal expression of

Igf2 includes face muscle and tongue, the connective tissue and smooth muscle cells of the viscera and blood vessels, and in both the epicardium and myocardium of the heart.

Igf2 is expressed widely in tissues derived of endodermal origin [95]. In the lung, *Igf2* transcripts can be detected in the cuboidal epithelia surrounding the bronchioles. High levels of expression are observed in the liver from the earliest stages of its development, and in humans this expression persists into adulthood [96]. Other sites of *Igf2* expression in endodermally derived tissues include the gastrointestinal tract epithelium, pancreas, salivary gland thyroid and thymus [95].

Stylianpoulou et. al. [97] observed a strong hybridisation signal for rat *Igf2* mRNA localised within each of the four cerebral hemispheres of the adult rat brain. On closer examination this staining pattern was found to be localised to choroid plexus epithelial cells and possibly the stromal blood vessels within; and the pia mater and arachnoid mater, the leptomeningial layers lining the cerebral hemispheres, the cerebellum and the brain stem. In mouse late-gestation embryos (e14-18) *Igf2* mRNA was found at high levels in the choroid plexus epithelium and stroma, as well as in the pia mater and arachnoid mater of the meninges [98], suggesting that *Igf2* expression in rat and mouse are directly comparable. The brain-specific activity of *Igf2* appears to be mainly confined to the exchange tissues during development, though expression has been reported in other structures arising from the neurepithelium, in both rat [95], and mouse [99]. These tissues include the fetal otic vesicle, Rathke's pouch, and in the postnatal mouse cerebellum, localised to the parenchyma and peaking at 3-4 days after birth [100]. Adult expression of *Igf2* is confined to the choroid plexus and leptomeninges in rodents [97], with the addition of liver expression in man [101], [96]. Expression of this gene in the remaining tissues declines shortly after birth [95].

The onset of expression of the *H19* gene lags slightly behind that of *Igf2*, and is first observed in the embryo proper at e8.5, in the developing mesoderm. In later development, *H19* transcripts are broadly distributed in the derivatives of the mesodermal and endodermal germ layers, such as in the liver, cartilage, intestine, heart, tongue, mesenchyme, kidney and the urogenital ridges. The most significant expression of *H19* is seen in those tissues of mesodermal origin, where the transcript is very abundant. While *H19* is transcribed in cartilage, like *Igf2* no transcripts are detected in the ossification centres [93]. *H19* is expressed in the choroid plexus and leptomeninges of the brain in

rodents [98], [102], but not in man [94]. In the adult, *H19* continues to be expressed in skeletal muscle (though at approximately 10-fold lower levels than in foetal muscle), whereas in all other tissues the expression of this gene cannot be detected beyond a few days post partum [93].

Cis-elements responsible for Igf2/H19 gene expression

While the expression patterns of the *Igf2* and *H19* genes during embryonic and post-embryonic development has been extensively studied, in a range of mammalian species, the identification of enhancer elements responsible for these expression patterns have not yet been fully resolved. The only enhancer elements demonstrated at present to drive tissue-specific expression at the *Igf2/H19* locus were isolated in a region at +9 to +11kb relative to the start of *H19* transcription [103], designated here as the *H19* enhancers.

Yoo-Warren et. al. [103] used an *in-vitro* system to define sequences downstream of the *H19* gene that were able to drive gene expression above basal levels in a cell culture system. This 3kb region was found to be able to significantly upregulate the *H19* promoter and the *tk* promoter in the transformed liver cell line *Hep3B*. The region exerted this activity in either orientation, one of the classical definitions of enhancer sequences. Further characterisation revealed two blocks of enhancer activity, with an additive effect on gene expression. These enhancers demonstrated tissue specificity, i.e., the upregulation of reporter gene expression could not be recapitulated in *HeLa* cells, which do not express *H19*. The endoderm specificity of these elements was tested by transient transfection into *PC13* cells. This murine embryonal carcinoma line differentiates into visceral endoderm in the presence of retinoic acid. Low levels of tagged *H19* expression was observed when the construct was transfected into undifferentiated *PC13* cells. This level was substantially increased when the cells were differentiated. The *H19* enhancers therefore demonstrated tissue- and developmental stage- specificity.

The *in vivo* function of these enhancer elements was clarified by a germline knockout of their endogenous location [104]. On maternal transmission of the deletion, *H19* expression was severely downregulated in a subset of embryonic and neonatal tissues. In the reciprocal experiment, when the deletion was inherited from the paternal allele, an apparently identical pattern of tissue-specific down-regulation of the *Igf2* gene was observed. Thus the *H19* enhancers were necessary for upregulation of both *H19* and *Igf2 in cis,* in a subset of tissues.

The study attempted to quantitate the extent of the loss of gene expression following deletion of the enhancers, and found a pattern consistent with the *in-vitro* characterisation of the *H19* enhancers. The reduction in gene expression of either *Igf2* or *H19* correlated roughly with the proportion of endodermally derived tissue present in the sample. In neonatal liver samples, which are thought to be almost entirely derived from the endoderm, gene expression was reduced to approximately 10% of wild-type levels when the deletion was *in-cis*. In gut, kidney and lung, where the endodermal component of the tissue is thought to be lower, 25-30% of wild-type expression levels remained. *In-situ* hybridisation analysis for *Igf2* and *H19* mRNA on e13.5 embryos confirmed that expression of these genes was only affected in cells of endodermal origin. For example, in gut wild-type animals express *H19* RNA at high levels in the gastric epithelial cells that line the lumen, and in the smooth muscle that surrounds it. When the enhancer deletion was transmitted maternally, only the smooth muscle expression of *H19* was retained.

This study also observed that there was very little decline in gene expression in skeletal muscle, brain, yolk sac or placenta following deletion of the enhancers.

A transgene composed of an internally deleted *H19* gene and the *H19* enhancers [105] expressed a tagged *H19* gene in neonatal liver, gut and yolk sac. These findings are in contradiction with the previous study, where yolk-sac expression was not lost following the enhancer deletion. This could be explained by the groups' analysis of different developmental stages, or that the presence of reporter gene expression is more easily detectable than a partial absence of gene activity, which might even be compensated for by other elements.

Mice that carried a transgene construct containing the *Igf2* promoter 2 to promoter 3 region, fused to a *LacZ* reporter gene and driven by the *H19* enhancers [106] reproducibly expressed β-galactosidase in liver, the epithelial layer of the gut, sclerotome, neural tube and the placodes of the cranial nerves of e11.5 embryos.

While the endoderm-specific activity of the *H19* enhancers has been well demonstrated, it is clear from the studies cited above that these enhancers are not sufficient to account for the abundant activity of *Igf2* and *H19* in non-endodermal tissues, but may have some activity there.

While a single element may not be responsible for the entirety of *Igf2* (and perhaps *H19*) expression in these mesodermally derived tissues, no *Igf2/H19 cis* sequences to date have been demonstrated to be mesodermal enhancers. Such elements are thought to lie *in-*

cis to the genes that they control, as a 130kb transgene [36], [107], containing sequences from the P1 promoter of *Igf2* to 35kb downstream of *H19* was expressed in many mesodermal cell types, such as skeletal muscle and tongue. This study confirms that there is no single mesodermal enhancer, as the heart expression of *Igf2* was not demonstrated by this transgene.

The role of the *H19* enhancers in the brain-specific expression of *Igf2* is unclear, as some *H19* enhancer-driven transgenes appear to express reporter genes in the exchange tissues [108], [109] whereas others do not [106]. As mentioned above, the deletion of these enhancers *in-cis* does not appear to diminish the expression of *Igf2* in brain tissues.

Tissue specificity of Igf2/H19 imprinting

It is a common feature of imprinted genes that their imprinting is not maintained in all tissues at all developmental time points. *Igf2* and *H19* follow this pattern, and show a dynamic pattern of imprinting throughout development.

The *Igf2* gene is biallelically expressed in the leptomeninges and choroid plexus in rodents [70], [102] and in man [101], and in rodents these tissues provide the only sites of *Igf2* expression after birth. In the rat, the expression of *Igf2* in the choroid plexus is at first monoallelic, (transcription from the paternal allele only was detected at days 13.5 and 15.5 of gestation, when the tissue is well differentiated); with a subsequent switch to biallelic expression by day 18.5. Thereafter equivalent expression from both alleles was evident until at least three months of age [102]. Biallelic expression of *Igf2* was first reported at e16 in the mouse foetus [70], which corresponds to approximately day 18 in rat development. Whether the same developmental switch in allele specific expression of *Igf2* occurs in the mouse is unclear, as earlier developmental stages have not been analysed. The imprinting of the *H19* gene is preserved in the choroid plexus and leptomeninges in the mouse [98] and rat [102], but not in humans where *H19* is silent [101].

In humans *Igf2* is expressed postnatally, and biallelically in the liver. The adult liver transcript of *Igf2* is transcribed from the P1 promoter which is human-specific [96]. The P1 promoter is always expressed biallelically [76]. While the expression of *Igf2* in the exchange tissues also occurs postnatally, the mechanism by which the gene escapes imprinting in these tissues appears to differ from the mechanism in human liver. The adult liver transcript in humans is strictly P1 derived, whereas exchange tissue-specific

expression originates from all four human promoters, with a very minimal contribution from P1 [101]. Additionally, an equivalent promoter to P1 has not been found in rodents.

The *Igf2* gene is expressed with a paternal bias from its earliest expression (at low levels in post-zygotic activation 4-cell embryos) [86]. However, expression from the maternal allele of *Igf2* has been reported in blastocysts [86], and in embryonic stem cells (ES cells, [110]), but expression levels were generally very low at these stages [86]. In later stages of development, *Igf2* mRNA levels from the maternal allele have been predicted at approximately 5% of total *Igf2* mRNA [8]; though one report [111] has estimated the level of maternal transcripts to be as high as 25% of the total expression of *Igf2,* in foetal liver cells at e13.5.

For *H19*, expression appears to be biallelic from its onset at implantation stages [86], with equal transcription from both alleles in embryonic and extraembryonic tissues. ES cells have also been shown to express *H19* biallelically [112], [110], though the expression of the paternal allele has been reported to be the result of prolonged culture of embryos *in vitro* [113].

Expression from the paternal *H19* allele has been reported to persist in embryonic and amniotic tissues until e10.5, and in the extraembryonic tissues up to at least e16.5 [86]. It is of interest to note that the relaxation in the imprinting status of *H19* in cultured embryos [113] was confined to the extraembryonic tissues, particularly the trophectoderm. In human first trimester placentas, paternal *H19* transcripts can be detected in the extravillous cytotrophoblast cells [114]. The extravillous cytotrophoblast component of Complete Hydatidiform Moles, which represent human androgenetic conceptuses, also express *H19* [115]. In the villous cytotrophoblast, from which the extravillous cells are derived, *H19* is expressed only from the maternal allele, suggesting that *H19* imprinting is relaxed in this tissue as development proceeds. A similar situation is observed in the mouse, where the primary giant cells of the e7 placenta have been shown to express the paternal allele of *H19* at high levels in a mosaic pattern. Transcription of the paternal allele in these cells progressively diminishes as development proceeds [116].

While reports are contradictory, the evidence seems to favour an initial phase of biallelic expression of *H19*, which either persists, or is re-introduced in a tissue-specific manner in later development. Transcription of the paternal allele of *H19* is confined to the extraembryonic tissues in later development, in both the mouse and in humans.

Conversely, while a low level of maternal *Igf2* expression can be detected at all stages of development, a strong bias toward paternal allele transcription is evident from the earliest stages of expression of this gene.

Epigenetic modifications

An imprinting mark, i.e., that primary modification of DNA that distinguishes the parental alleles of a gene, must conform to certain criteria:

i) It must be a reversible modification of DNA, which can be erased in the germline of the parent, then reset according to that parent's sex.

ii) The mark must be stable throughout subsequent cell divisions, being inherited from one chromosome to its daughter chromosomes.

iii) It must direct the transcriptional state of the gene that it controls.

DNA methylation

DNA methylation in mammals occurs predominantly at the 5' position of Cytosine residues in CpG doublets. This modification of DNA is clonally heritable, as the recognition site for the major mammalian maintenance methylase, *Dnmt1* is palindromic, and the preferred substrate for this enzyme is hemi-methylated DNA (reviewed in [117]).

CpG methylation is important for development as mouse embryos homozygous for a partially inactivated allele of *Dnmt1* die before e10.5, and a null allele causes lethality at the 5-somite stage. ES cells homozygous for either of the alleles proliferate normally, with their genomic DNA highly demethylated, but die upon the induction of differentiation [118]. The expression of imprinted genes is disrupted in mice lacking functional *Dnmt1*, for example, the *H19* gene becomes activated on both parental alleles whereas *Igf2* is silenced [119]. The inappropriate expression of imprinted genes cannot in all cases be restored by reintroducing a functional copy of *Dnmt1* at postimplantation stages. Embryos derived from ES cells expressing an integrated *Dnmt1* cDNA in a *Dnmt1*$^{-/-}$ background do not regain allele specific expression of imprinted genes, despite remethylation of bulk genomic DNA [120].

The experiments in which *Dnmt1* is inactivated in mouse development clearly show that methylation plays a role in the imprinting mechanism. The case proposing CpG methylation as an imprinting mark is strengthened by the observation that regions within or

near most imprinted genes are methylated in a parent of origin dependent manner (reviewed in [21])

CpG methylation levels of DNA are highly dynamic during development, and male and female gametes differ markedly with respect to global levels of DNA methylation ([121]; and reviewed [122]). In the male germ-line, germ cells are relatively unmethylated, but maturing gametes acquire methylation as spermatogenesis progresses. Mature sperm have very high levels of methylated CpGs. Conversely, mature oocytes attain a much lower level of genome methylation during their development. At fertilisation, and until the late 2-cell stage, paternal and maternal chromatin can be distinguished on the basis of their methylation status. Immediately following fertilisation, paternally derived chromatin is highly methylated, but during the first cell division the paternal genome is rapidly demethylated [21], [123], presumably due to the action of a demethylase protein [124]. The maternal genome shows moderate levels of methylation at fertilisation, and these levels decline during the first few cell divisions at a rate suggestive of a loss of maintenance methylation, rather than the active demethylation observed in paternal chromatin [21], [123]. By the blastocyst stage of development both genomes are grossly indistinguishable, and almost completely demethylated. Remethylation of the genome is initiated after implantation, and levels slowly rise during subsequent development. Extraembryonic tissues maintain lower levels of global genome methylation than embryonic tissues throughout development, and germ cells remain unmethylated [121].

During early development the maternal and paternal genomes are clearly distinguishable in terms of their methylation status, thus CpG methylation provides a good candidate for a parental allele specific imprint. This imprint must persist, however, throughout the genome wide changes in methylation status in the early embryo.

Methylation also provides a candidate for the maintenance of a silent transcriptional state found on the inactive allele of imprinted genes. There is a general correlation between methylation of promoters, and gene inactivation [125]. Though the genome is globally remethylated during implantation stages, the CpG islands associated with housekeeping genes escape modification, and at later stages of embryogenesis the promoters of tissue specific genes are demethylated in the cell types in which they are expressed (reviewed in [126]). CpG methylation is also found on the inactive allele of the X

chromosome [122]. Furthermore, microinjection of methylated gene sequences into *in vitro* cell culture systems results in the formation of inactive chromatin (reviewed in [127]).

Methylation at CpG residues therefore provides many of the necessary qualities of an imprinting mark; it can distinguish the maternally and paternally derived genomes, can be stably inherited at cell division, and can mediate changes in gene activity.

Methylated regions at Igf2/H19

DNA methylation was shown to be involved with imprinting at the *Igf2/H19* locus by analysis of mice containing deletions in the DNA methyltransferase gene *Dnmt1* [119]. When homozygous mutant embryos were examined for expression of *H19* and *Igf2*, the normally silent paternal *H19* allele was found to be activated, whereas the expression of *Igf2* was repressed. Thus the loss of DNA methylation had opposite effects on the two genes.

Several regions within the locus including the *Igf2* and *H19* genes differ in respect to the methylation status between their parental alleles. In order for a region of differential methylation to qualify as an imprinting mark, the appropriate residues must be methylated in the maternal or paternal genome while the two genomes are separated. This separation extends from the onset of gametogenesis until syngamy of male and female pronuclei, just prior to the first cell division of the 1-cell embryo. The modification must then persist throughout development, i.e., one allele must resist the genome-wide demethylation in the blastocyst, and the other allele must resist the *de-novo* methylation at implantation stage.

Three differentially methylated regions (DMR*s*) have been identified close to, or within the *Igf2* gene, and are depicted in ***Figure 2a***.

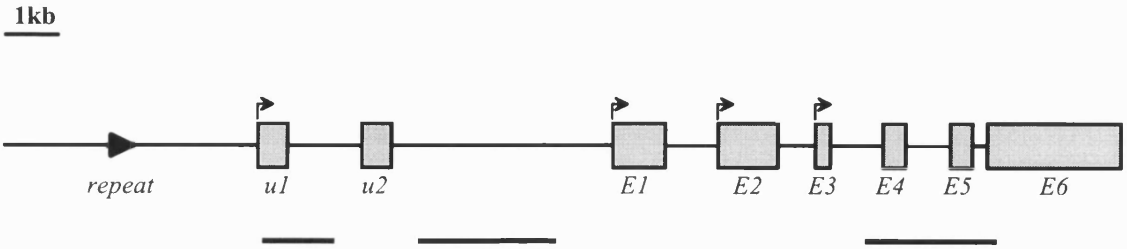
DMR0 [6], is immediately 5' to *Igf2* upstream exon 1. This region exhibits maternal allele-specific methylation in placental tissue, though both alleles are methylated in the foetus. This region is associated with the start of transcription of both a sense [6] and an antisense [6, 14] *Igf2* transcript, which are both expressed in the placenta from the paternal allele.

DMR1, an element located between upstream-exon2 and exon 1, was first discovered to be differentially methylated by Sasaki et. al. [8], who compared the digestion of this region by methylation-sensitive enzymes between normal embryos, and MatDi7 embryos (those with two copies of maternal chromosome 7, and no paternal chromosome 7). MatDi7 embryos showed a pattern of complete digestion in this region, indicating that

maternal DNA is unmethylated, but normal embryos showed some protected fragments which must correspond to paternal, methylated residues. Brandeis et. al. [128] further refined this region to ~600bp approximately 3kb upstream of *Igf2* promoter1 (P1). Using inter-specific crosses of *Mus musculus* and *Mus spretus* in order to identify parental alleles by sequence polymorphisms, four CpGs were identified which were partially methylated on the paternal allele and completely unmethylated on the maternal allele in adult liver and embryos at e15-e18. In sperm and oocytes, all four sites were methylated. In the 16-32 cell morula, this region was unmethylated on both alleles except for one site (site 3), which was more highly methylated on the paternal allele. The differential methylation pattern observed in somatic cells must be established during post-fertilisation development, which disqualifies this region as a site for the primary imprinting mark.

A systematic study of the methylation status of all of the CpGs in this region was carried out using the bisulphite sequencing technique. Feil et. al. [7] found a gradient of methylation over 13CpG dinucleotides in this region, from almost complete methylation of CpG 1 (near to upstream exon 2) to virtually no methylation at CpG 13 (most proximal to P1). Overall, maternal chromosomes were less methylated than paternal chromosomes (23% and 38% respectively), though no single CpG was always methylated paternally and unmethylated maternally. For CpGs 1 to 4 (of which site 3 above is one) the differences between parental alleles was most pronounced, with ~50% more paternal methylation than maternal-specific methylation. In adult choroid plexus this region is highly methylated on both alleles.

DMR2 was initially described as three differentially methylated HpaII restriction sites, one within intron 4 directly upstream of exon 5, one within exon 5, and one located within the coding sequences of exon 6. These sites were found to be more methylated on the expressed paternal allele than on the repressed maternal allele. In fetal brain this region is almost entirely unmethylated, while in fetal liver there is a high level of paternal methylation. In adult choroid plexus this region is methylated on both chromosomes. The pattern of methylation of this region is thought to correspond to the expression status of *Igf2* [7]. DMR2 is conserved in horse and man, and in horse and mouse it has been shown to form a stable stem-loop structure. The equine stem-loop motif was shown to bind a specific protein fraction from neonatal liver nuclear extracts, of both horse and mouse, in gel-mobility shift assays. Protein binding was reduced upon methylation of the CpGs in the stem loop area [129].



	DMR0	*DMR1*		*DMR2*	
sperm	nd	++		++	
oocytes	nd	++		nd	
paternal blastocyst ***maternal***	nd	+		nd	
paternal adult ***maternal***	- / +	+ / -		+ / -	

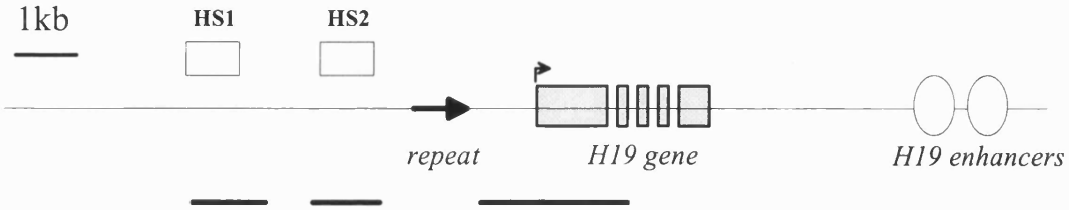
Figure 2a. Schematic representation of the *Igf2* gene region showing regions of allele specific methylation (black horizontal bars, ***DMR 0, 1, 2***). The eight exons of *Igf2* are depicted as grey boxes (u1, u2, and E1-6).The four promoters are depicted as arrows above the exons.

The table below indicates at which points in development these regions are methylated (++, highly methylated; +, methylated; - , unmethylated, nd represents those time points where information is not available). A detailed description of this region, with the corresponding references is contained within the text.

The *H19* gene region also contains several regions of differential methylation, which are depicted in ***Figure 2b***.

The H19 gene and promoter-proximal region. Ferguson-Smith et. al. [9] first identified two *HpaII* sites within the first exon of the *H19* gene which were fully unmethylated in MatDi7 embryos, 50% digested in normal embryos. They found the remainder of the gene to exhibit the same methylation pattern in normal and MatDi7 embryos, notably at the 3' end of the gene near the enhancers. This analysis was extended ([38], [128]) to include further CpG residues within the gene. These studies revealed that *H19* is modified in the first exon of the gene on the paternal allele regardless of expression status, being methylated in the brain, adult liver and in embryonic stem cells where *H19* is known not to be expressed. An exception to this is in the adult choroid plexus where both alleles of the *H19* gene are fully methylated [7]. Sperm DNA is highly methylated in this region, but there is no differential methylation during morula and blastula stages of development, which rules out the *H19* gene methylation as a candidate for the primary imprinting mark ([38], [34], [130]).

The *H19* promoter contains a CpG island which is highly methylated on the paternal allele, and undermethylated on the maternal allele [9]. This promoter region acts in a manner typical of CpG islands in that it is unmethylated in sperm and in blastocysts, [9], [128], [38]), and is remethylated in the genome-wide *de-novo* methylation in postimplantation development [113], [121]. Methylation of the promoter of the paternal *H19* allele appears to correlate closely with *H19* gene silencing. This region is progressively methylated in the embryo during postimplantation development [128], [113], corresponding to the gradual silencing of the paternal *H19* allele during this phase of embryogenesis [86]. The progression of methylation at the paternal *H19* promoter is more protracted in the extraembryonic lineages, especially in the trophoblast [113], where expression from the paternal *H19* allele is most pronounced [86], [116]. Furthermore, this region is methylated on the paternal allele in the choroid plexus [7], demonstrating that the methylation of this region corresponds to the imprinting status of the *H19* gene, and not that of *Igf2*. While the lack of methylation of this region in early embryogenesis disqualifies the *H19* promoter as a candidate imprinting mark, its methylation closely parallels the silencing of the paternal *H19* allele, and gives an indication that the marking and silencing of imprinted genes may be separable processes.



	DMDA	*DMDB*	*H19 promoter and gene*	
sperm	+++	+++	-	+++
oocytes	-	-	-	
paternal	++	++		
blastocyst			-	
maternal	-	-		
paternal	++	++	++	
adult				
maternal	-	-	-	

Figure 2b. Schematic representation of the *H19* gene region showing regions of allele specific methylation (black horizontal bars, ***DMDA, DMDB,*** the ***H19 promoter and gene***), and the regions that are hypersensitive to nucleases on the maternal allele (**HS1** and **HS2**, white boxes).

The table below indicates at which points in development these regions are methylated (++, highly methylated; +, methylated; - , unmethylated, nd represents those time points where information is not available). A detailed description of this region, with the corresponding references is contained within the text.

H19 5' region (DMD). The region encompassing 10kb upstream of the *H19* promoter shows all of the features of an imprinting mark. This region is heavily methylated on the paternal allele in neonatal mice, and in sperm DNA [38, 34, 130]. A more detailed analysis of this region reveals two blocks of CpG dinucleotides which show paternal specific methylation in neonatal and adult tissues, as well as in pre-implantation development and in sperm [131]. These two blocks; DMDA (which spans from –4.1 to –3.8kb of *H19* transcription start, see ***Figure 2b***) contains 11 CpG dinucleotides which were shown by bisulphite sequencing to be highly methylated in sperm, almost entirely unmethylated in mature oocytes, and 50% methylated from the fertilised egg through to adult tissues. This region contains the HpaII and HhaI restriction sites initially shown by Tremblay et. al. [34] to fulfil the criteria of an imprinting mark. Bisulphite sequencing of 24 CpGs in DMDB (from –2.6 to –2.1kb of *H19* transcription start, see ***Figure 2b***) shows that this region is highly methylated in sperm, unmethylated in oocytes, and throughout early development the paternal allele is more highly methylated than the maternal allele. There is a low level of maternal methylation that is again consistent with the methylation pattern of *HpaII* sites found by Tremblay et. al. [34] in this region.

Resetting the imprint

Imprinted genes are thought to re-establish their sex-specific imprint during gametogenesis, as at these stages the male and female genomes are separated. Many studies have attempted to correlate methylation changes at imprinted loci to a sex-specific difference in germ cell potency. If methylation is the 'imprinting mark', i.e. the epigenetic modification which is causal in the establishment of differential gene expression between parental alleles, then certain predictions can be made about the temporal changes in methylation status at imprinted genes as development proceeds. It would be predicted that in germ cell development the methylation patterns of the previous generation should be erased, and then replaced with a methylation pattern specific to the sex of that germ cell. In support of this, it is known that primordial germ cells (PGCs) are markedly undermethylated globally, as compared to other embryonic tissues, from at least e11.5 [121].

As discussed above the *H19* DMD displays all of the characteristics of an 'imprinting mark', i.e., this region is differentially methylated in mature gametes, and the differences in methylation status persist throughout subsequent development.

What is the time-scale of the establishment of paternal specific methylation at the *H19* locus? In contrast to many other imprinted genes [132], the somatic pattern of *H19* methylation is not lost until late in PGC development. In cells derived from e12.5 PGCs, CpG residues in the *H19* DMD and promoter-proximal region had the somatic pattern of methylation, i.e., the paternal allele was hypermethylated and the maternal allele was hypomethylated. By e14, this somatic methylation pattern is lost, as judged by an experiment in which the nucleus from e14-16 PGCs were transferred into enucleated oocytes [133]. In embryos derived from these cells, *H19* is biallelically expressed at e9.5 (the latest stage of embryo recovery), suggesting that the correct paternal-specific modification was never established, and the gene was unmethylated at the DMD. Bisulphite sequencing of the DMD and promoter-proximal regions of the *H19* gene in male PGCs [134] at e13.5 revealed that these regions were undermethylated on both parental alleles. If the parental imprint had been erased by this stage (e13-e14), both alleles should be equivalent. It appears, however, that differences between parental genomes persist until late into spermatogenesis, with both alleles of *H19* becoming truly equivalent in meiotic spermatocytes [135]. In mitotic spermatogonia the maternal allele of *H19* is significantly less methylated than the paternal allele at the DMD and promoter-proximal region. During the meiotic stages of spermatogenesis the maternal allele comes to resemble the paternal allele, until the round spermatid stage of post-meiotic spermatogenesis, when the origin of the two alleles cannot be distinguished by their methylation status.

It appears then that though the two parental alleles become equivalent in terms of their demethylation in embryonic PGCs, differences still exist between the alleles, which affect the ability of each to acquire the sperm-specific methylation pattern. Further CpG residues could exist outside the regions examined, that retain their paternal methylation pattern and thus distinguish the parental alleles; or the two alleles could be identical in terms of their undermethylated state in late gestation PGCs, but other epigenetic modifications could still exist.

Analysis of the expression of both the *Igf2* and *H19* genes during gametogenesis and in early embryogenesis suggests that the expression of these genes is decoupled from methylation at the DMD. *Igf2* and *H19* are biallelically expressed (at low levels) in meiotic germ cells from both sexes [136], in pre-implantation embryos [86], and in ES cells [112], when differences in the methylation status between parental alleles has been demonstrated.

Chromatin modification

The simplest repeating structure of chromatin is the nucleosome, which consists of ~200bp of DNA wrapped approximately 2 turns around an octomer of the highly conserved core histones H2A, H2B, H3 and H4.

The arrangement of DNA into nucleosomes is thought to have a generally repressive effect upon transcription, and this repression must then be overcome at specific loci by positive acting transcriptional regulatory proteins. In yeast, inhibition of histone synthesis leads to constitutive activation of many genes, even in the absence of their specific activator proteins (reviewed in [137]).

Alterations in chromatin structure can be observed at sites of active genes, prior to the onset of transcription, which can be visualised by the altered sensitivity of such regions to cleavage by nucleases (the endonuclease DNase I is most commonly used), and termed DNase sensitive regions. Early studies of the avian globin, lysozyme and ovalbumin loci revealed that DNase sensitivity extends across these loci in cells which are competent to transcribe the genes (reviewed in [138]). In contrast, heterochromatic regions of the genome, where transcription is repressed, are extremely insensitive to nucleases.

Sensitivity to nucleases delineates domains in which transcription of genes occurs. Within these regions, sequences that show an even greater sensitivity to nucleases (nuclease hypersensitive sites, HS) often correspond to promoters, transcription factor binding sites, and enhancers [137]. Like the generalised nuclease sensitivity in active gene regions, the formation of HSs at the regulatory regions of genes appears to be a prerequisite for transcriptional initiation, and represents an early step in gene activation.

Regions of active transcription with increased sensitivity to nucleases often correspond to regions with high levels of histone acetylation. Acetylation occurs at Lysine residues of the amino-terminal 'tails' of the core histones. The acetylation status of histone H4 has been most extensively studied. H4 can be acetylated at four Lysine residues at the amino-terminus, and in mammals the order in which these residues are modified occurs in a fixed order. Lysine 16 is modified first, followed by Lysines 12 or 8, and finally Lysine 5. Acetylation of Lysine 5 therefore can be used as a marker for hyperacetylated histones [139].

The histone acetylation status of localised chromatin domains provides a mechanism for the propagation of imprinting marks. Though replication disrupts nucleosome association, the disruption is transient and the histones originally associated

with a region of replicating DNA will remain associated with that region and reassemble after the passing of the replication fork. Modifications of histone proteins will therefore remain in contact with their particular genes and are stably propagated within their local environments (reviewed in [140]).

The acetylation status of histones is highly dynamic during preimplantation development. Sperm chromatin is packaged with protamines rather than histones. Immediately upon sperm entry the paternally derived chromatin sheds the protamines, and replaces them with histones. The source of these histones is the oocyte cytoplasm, which contains a large store of acetylated histones. The maternally-derived chromatin is relatively underacetylated in oocytes, and remains so until the onset of the first round of DNA synthesis, just prior to syngamy [141].

The male chromatin aquires a high proportion of acetylated H4 after fertilisation, but the female chromatin does not, unless the oocyte is artificially activated. In parthenogenetic embryos, all chromatin aquires high levels of acetylated H4. There is a high turnover of histone acetylation and deacetylation at these early developmental stages. Inhibition of histone deacetylases by Trichostatin A (TSA) results in female chromatin becoming hyperacetylated. These observations suggest that the dynamics of acquisition and maintenance of histone acetylation differ between male and female chromatin in the zygote [141]. Such differences could provide the basis of a mark that distinguishes the alleles of imprinted genes.

Acetylation of H4, in general, occurs in regions known to be enriched in coding DNA, and furthermore, chromatin fractions enriched for sequences characteristic of CpG islands also contain highly acetylated histones. Conversely, H4 in heterochromatic regions is consistently underacetylated, for example in centric heterochromatin [139], and on the inactive X chromosome of mouse and man [142].

How does histone acetylation affect gene expression? Acetylation of histones is thought to neutralise the positive charge of the Lysine tails, and thereby reduce the affinity of the histone octomer for DNA. Acetylation therefore provides a mechanism by which DNA can be rendered generally accessible to *trans*- acting factors while still maintaining a nucleosomal structure. Transcriptional control requires the binding of multiple sequence-specific transcription factors to regulatory elements *in cis* to the transcriptional start site. Activation domains within these factors recruit co-activators, which, in turn, facilitate the activity of the basal transcriptional machinery (reviewed in [143]). Proteins with histone

acetylase activity have been identified, which previously have been characterised as components of the RNA Polymerase II (Pol II) transcription machinery, as proteins that associate with transcription factors, and proteins that positively affect transcription *in vivo* (reviewed by Struhl, 1998). Examples of this are: the TAF130/150 histone acetylase, which is a subunit of the TFIID complex, a basic component of the Pol II transcriptional machinery in all eukaryotes; p300/CBP histone acetylase was described initially as a transcriptional co-activator that functions by interacting with a wide variety of enhancer binding proteins, including the glucocorticoid receptor.

The best described histone deacetylases are members of a gene family that includes the human deacetylase HDAC1, and yeast Rpd3. HDAC1 and Rpd3 are found in large multiprotein complexes that include the Sin3 co-repressors, and other negative regulatory proteins such as Mad and NCoR [144].

Recently a connection has been discovered between methylation at CpG residues and gene silencing by the mSin3A/NCoR/HDAC1 complex, mediated by the methyl-binding protein MeCP2 [145, 146]. MeCP2 is a member of a family of proteins containing a methyl CpG-binding domain (MBD), (reviewed in [147]). MeCP2 binds preferentially to methylated CpG doublets via the MBD, in any sequence context. In general, there are approximately $5x10^6$ molecules of MeCP2 per nucleus in the mouse, corresponding to one molecule per 5-20 nucleosomes [148], indicating that MeCP2 is an important structural component of chromatin. MeCP2 contains a transcriptional repression domain (TRD) that can confer silencing upon reporter constructs when fused to a GAL4 DNA-binding domain, and can operate over several hundred base pairs to silence transcription. The transcriptional repression can be alleviated by addition of the histone deacetylase inhibitor TSA [147]. It is the TRD which interacts directly with mSin3A, and thereby recruit histone deacetylases to DNA marked by methylation [145]. Methylation of critical regulatory regions of imprinted genes may act to recruit chromatin remodelling factors in an allele-specific manner, leading to the downregulation of target genes.

Proteins which modify the histone acetylation status of chromatin therefore appear to act both generally upon the basal transcriptional machinery, and can be recruited to specific promoters/chromosomal regions, bringing about local changes in chromatin conformation which directly affect the transcriptional status of specific genes. As with MeCP2, the recruitment of chromatin remodelling factors may require the marking of suitable sequences by CpG methylation, but this is by no means the only mechanism, as

the mSin3/NCoR/HDAC1 complex can also act to silence genes by a methylation independent mechanism [145].

CpG methylation does not provide the only candidate for an epigenetic mark that is stable through DNA replication. The proteins involved in the formation of open or closed chromatin are intimately connected to DNA, and are very likely to co-segregate with their respective binding sequences during chromosome replication and division. Chromatin proteins are known to recruit other members of their multiprotein complexes, therefore fully competent chromatin complexes can be formed from a small number of components after cell division. In this manner, the association of specific protein complexes with DNA could constitute an epigenetic mark, which is heritable through multiple cell divisions. In *Drosophila melanogaster*, where there is no genome methylation to provide epigenetic marks, heterochromatin at centromeres and telomeres is stably inherited [149]. Repressive complexes at the homeotic genes, which are mediated by the Polycomb group proteins are stably inherited, by a mechanism which does not require the formation of canonical heterochromatin (reviewed in [40, 150]).

Chromatin structure at Igf2/H19

Several regions of nuclease HS have been identified at the *Igf2/H19* locus (see ***Figure 1***), and the acetylation status of the histones has also been ascertained in some cases, though the developmental profile for the formation of such regions is less well defined than the regions of differential methylation.

In the mouse the *Igf2* gene region displays a number of HSs, the strongest within the CpG island at P2, with others within exon 2, at P3 and within exon 3 [8]. These HSs can be found in tissues that express *Igf2*, but not in adult spleen, which does not express the gene. A comparison of DNase I HSs at these sites between MatDi7 and normal embryos reveals that there are no allele specific differences in chromatin compaction within the *Igf2* coding region, and that both alleles are equally accessible to nuclear factors. The region encompassing DMR0 displays a constitutive HS, which may mark the promoter of the placental-specific transcript of *Igf2* [6]. The region including DMR1 has been found to be accessible to nucleases [151], with no difference in sensitivity between parental alleles. Feil et. al. [152] found four HSs partially overlapping the DMR1 region, which were equally DNase I hypersensitive in androgenetic and parthenogenetic ES cells, as well as in embryonic chromatin. These HSs do not persist in neonatal liver, however, which suggests

that nuclease sensitivity within this region does not generally reflect *Igf2* expression. Feil et. al. [152] have suggested that the HSs at DMR1 may correlate with the expression of the antisense transcripts [14, 6] that are transcribed from the upstream region of *Igf2*. The DMR2 region does not display enhanced sensitivity to nucleases in any tissues assayed to date [151, 152].

In a systematic analysis of 130kb including *Igf2* and *H19* for regions of nuclease sensitivity, a region approximately 32kb upstream of the *H19* start site was found which displayed a high level of nuclease sensitivity on both alleles. This region was hypomethylated with respect to the surrounding DNA, and found to be conserved in a range of mammalian species [151]. This region is the main subject of this report, and is designated the CCD (for centrally conserved domain).

The *H19* gene differs in respect to *Igf2* in that the inactive allele is insensitive to nucleases. The promoter region is marked by a strong HS, but only on the active, maternal allele [9]. There are also several HSs downstream of the coding region, three of which correspond to the location of the *H19* enhancers. These 3' HSs are all hypersensitive to nucleases on both parental alleles in chromatin derived from embryos at midgestation [9, 151].

The region approximately 2-4kb 5' to the *H19* transcriptional start site (the DMD) has been extensively studied in terms of its chromatin structure. Hark and Tilghman [153] discovered two blocks of nuclease hypersensitivity corresponding to two regions of extensive paternal CpG methylation (HS1 and HS2 in ***Figure 2b***). These HSs (at approximately -3.8kb and –2.5kb) are present only on the maternal allele, in tissues known to express *H19*, and in non-expressing tissues. A fine mapping study of this region [154] detected five HS in chromatin derived from adult brain, liver and kidney, and seven sites in ES cells. These HSs map within the regions of nuclease sensitivity described by Hark and Tilghman [153]. Furthermore, Khosla et. al. [154] treated chromatin derived from the tissues detailed above with micrococcal nuclease (MNase). MNase digests DNA preferentially at the linker regions between the nucleosomes. The most usual positioning of nucleosomes is in ~200bp arrays, and digestion of such DNA with MNase gives rise to a distinctive laddered digestion pattern. When the maternal 5' DMD region was subjected to this treatment a digestion pattern resulted that did not match that of a canonical nucleosome array, and the resultant pattern was suggestive of a chromatin structure in which the core

histones had been displaced. The areas of MNase sensitivity matched very closely to those determined previously as DNase I HSs. The paternal allele showed an identical MNase digestion pattern in this region to that of a non-imprinted gene, with the characteristic laddered pattern. The maternal MNase digestion pattern was also found in the DMD region in parthenogenetic ES cells, suggesting that the underlying chromatin conformation responsible for such a pattern is acquired early in development. Khosla et. al. [154] suggest that this 'unusual chromatin' conformation of maternal chromatin at the DMD reflects an opposite epigenetic state to the hypermethylated paternal allele, and may protect the maternal allele from *de-novo* methylation at implantation. Further work will be needed to determine if this region displays the same irregular chromatin structure in oocytes, and could thereby constitute a dominant epigenetic mark of the maternal genome. Kanduri et. al. [155] also reported maternal allele specific MNase hypersensitive sites within this region, two of which correspond to sites reported by Khosla et. al. [154]. In contrast to the previous report, however, the authors suggest that the MNase sensitive regions map to sites corresponding to the linker regions between histones, and that the previous evidence of a nucleosome free region was artefactual. Furthermore, these nuclease hypersensitive sites were found to overlap with the 21bp repeat elements identified previously as sites of sequence homology between the human, rat and mouse differentially methylated regions [156].

Further studies in the epigenetic status of this region have revealed that the promoter of the maternal allele of the *H19* gene is more readily bound by transcription factors in early development than the paternal allele. Transcription factors bound to promoter elements could outcompete nucleosomes and DNA methyltransferase, and thereby inhibit methylation, thus stabilising the expression differences between the alleles in later stages of development [157].

A 1.2kb fragment of the DMD containing HS2 has been shown to function as a silencer in *Drosophila*, suggesting a conserved mechanism in the silencing of the *H19* gene, and gene silencing in *Drosophila*. As *Drosophila* lacks genome methylation, this region may be recognised by different factors than in mouse (since it is the methylated form of this region which is linked to silencing of *H19*), to bring about a similar downstream event [158]. Alternatively, methylation of this region may not be related to its role in the imprinting mechanism.

Chromatin prepared from primary fibroblasts (which display correct imprinting of *Igf2* and *H19*) was immunoprecipitated with antibodies raised against the acetylated forms of histones H3 and H4. Maternal chromatin from the *H19* locus was found to be acetylated at much higher levels that paternal chromatin. *Igf2* is acetylated at comparable levels on both parental chromosomes [159]. This result confirms the earlier observation that the alleles of the *Igf2* gene are equivalent in terms of their potential transcriptional status [8], whereas the paternal allele of *H19* is silenced via a mechanism involving chromatin compaction [9]. There is some evidence that chromatin compaction of the inactive *H19* locus is most likely a gradual process. Firstly, the paternal allele of *H19* is known to be expressed at relatively high levels in early development [86], particularly in the extraembryonic tissues [86, 116]. Secondly, when e7.5 embryos were cultured *in-vitro* in the presence of the histone deacetylase inhibitor TSA, the paternal allele of *H19* was reactivated in a stochastic fashion (reminiscent of position effect variegation), giving rise to a small subpopulation of cells in the trophectoderm which expressed *H19* biallelically [116]. This experiment provides direct evidence that the silencing of paternal *H19* depends directly upon chromatin modifications, at least in the extraembryonic tissues.

In summary, studies of the epigenetic modifications of the *Igf2* and *H19* genes show that despite very similar spatial and temporal patterns of regulation, the mechanism mediating the imprinting of these two genes, while linked, may be quite different. The *Igf2* gene is transcribed with a paternal bias from the earliest stages of its expression, is in an 'active' chromatin conformation on both alleles (as judged by accessibility to nucleases, and hyperacetylation of histones H3 and H4), and is marked by methylation on the active, paternal chromosome. On the other hand, the *H19* gene displays biallelic transcription from its first expression, and expression levels remain high from the 'silent' paternal allele up to midgestational stages in the extraembryonic tissues. The inactive, paternal allele of *H19* is generally nuclease insensitive, methylated at the promoter, and within the 5' portion of the gene, and underacetylated.

Mechanisms of *Igf2/H19* imprinting

As was discussed above, a large number of structural studies have isolated regions *in cis* to the *Igf2* and *H19* genes that are distinctive in terms of their chromatin structure, and patterns of CpG methylation. While correlations can be found between, for example, levels of CpG methylation at a gene region and the expression levels of that gene, such analyses

do not causally link the two phenomena. In order to provide such a link, two major experimental strategies have been adopted in order to resolve the role of *cis*-acting elements in the mechanism/s of *Igf2* and *H19* imprinting.

The first of these strategies is the creation of transgenic mice bearing integrated constructs containing *cis* elements of the imprinting domain fused to reporter genes. An advantage of this approach is that, once found, active sequences can be characterised fairly easily. The disadvantages stem from the observation that the site of transgene integration can have a large effect upon the activity of the experimental construct, thus necessitating a comparison of several transgenic lines. In addition, the critical regions examined in transgene constructs may require a larger genome context in which to act.

The second approach is that of the generation of targeted germline deletions of candidate regions by homologous recombination, i.e., the creation of knockout mice. While the use of this approach ensures that the action of a particular element will be examined in an appropriate genome context, in a complex or redundant system compensating elements may mask the effects of removal of one element. Experiments utilising a transgenic approach therefore identify those regions, which are sufficient, for (in this case) the imprinting phenomenon, whereas experiments using a knockout approach identify those sequences which are necessary for the effect.

Elements required for H19 imprinting

Are any of the regions defined by nuclease HSs and/or differential methylation able to confer imprinting at ectopic loci? This question has been answered in part by a series of experiments involving transgenic mice bearing *H19* transgenes.

H19 transgenes which include 4kb upstream sequences, the *H19* structural gene, and 7kb downstream sequence are expressed when transmitted maternally, and silenced when transmitted through the paternal genome [37, 38, 39] if the transgenes are present at high copy number. Low copy number transgenes are always silenced. The transgenes contain much of the sequence of the 5' DMD CpGs, as well as the 3' *H19* enhancers. These imprinted transgenes are methylated on paternal CpGs in the 5' region as well as in the gene itself in a manner similar to the endogenous *H19* gene.

In transgenes that bear internal deletions removing the 3' part of exon 1, exon 2 and the 5' part of exon 3, the imprinting is maintained [37, 38]. Removal of the 2kb DMD at the 5' end of the construct leads to biallelic expression of the transgenes. A 3' truncation of the construct which removes the enhancers also results in loss of transgene imprinting, though

expression is also lost, and imprinting status is instead measured by transgene methylation. Deletion of a 451bp region containing the G-rich repeat element ~1.5kb upstream of *H19* start results in biallelic expression of the transgene, but curiously monoallelic expression is restored if 1.7kb additional 5' sequence is added to this deleted construct [156]. Replacement of the *H19* coding region with a *luciferase* reporter gene gives rise to biallelic expression of the transgene, as does a deletion which removes the first 700bp of the *H19* coding sequence (which includes the differentially methylated region of the *H19* gene) [39].

The manipulations described above demonstrate that imprinting of the *H19* gene at ectopic loci is probably a multifactorial process. That the imprinting and expression status of these transgenes is very delicate is implied by the need for many copies of the transgene, perhaps to insulate the region. Removal of i) the DMD, ii) the G-rich repeat, iii) exon 1 of *H19* and iv) the *H19* enhancers can all result in the loss of monoallelic expression, suggesting that many factors act co-operatively to give a stochastic pattern of expression.

This delicacy of the imprinting status of *H19* is indirectly supported by the observation that the only transgene created to date which gives rise to stable imprinting of *H19* at low copy number is a 130kb YAC [36]. It is also possible that the YAC transgenes contained regions required to stabilise *H19* imprinting which are not present in the smaller transgenes, but which can be functionally substituted for by multiple copies of those elements which are present.

The enhancer competition model

The observation of extremely similar temporal and spatial patterns of gene expression of *Igf2* and *H19* led to the idea that the two genes may be co-ordinately regulated, and that their imprinting might be mechanistically linked.

Deletion of the genomic region containing the *H19* enhancers demonstrated that the expression of the two genes is dependent upon a shared element, at least in a subset of tissues [104]. This led to the investigation of whether the imprinting mechanism too was shared by the two genes, and seemed to be confirmed by an experiment in which the *H19* gene, as well as approximately 7kb of flanking sequence was deleted [27]. When the deletion was transmitted from the male parent, both *H19* and *Igf2* gene expression was unaffected. Maternal transmission of the deletion resulted in activation of the *Igf2* allele *in cis* to the deletion. This experiment led to the proposal of the 'enhancer competition' model, which states that *Igf2* and *H19* are in competition for shared enhancers. On the

maternal chromosome, the *H19* 'wins', either due to its closer proximity to the enhancers, or due to superior promoter strength. On the paternal chromosome the *H19* gene is methylated, which leads to its inactivation and allows *Igf2* to be expressed. Support for this model came from the observation that the *Igf2* gene is in an open chromatin conformation, and therefore competent for transcription on both alleles [8], whereas the inactive allele of *H19* is insensitive to nucleases from early in development [9].

A necessary consequence of the enhancer competition model is that *Igf2* and *H19* cannot be expressed from the same chromosome. Mounting evidence suggested that the two genes could in fact be active *in cis*. Firstly, ES cells and early embryos in some cases expressed both *Igf2* and *H19* biallelically [113, 86]; and in the choroid plexus *Igf2* expression is biallelic and *H19* expression monoallelic [98]. Secondly, when the *H19* gene deleted and replaced by a *luciferase* reporter gene (with an intact *H19* promoter), paternal transmission of the new allele resulted in some cases in a high level of expression of *luciferase*. In the same individuals *Igf2* was expressed as normal, independent of the level of gene expression from the reporter gene on the same chromosome [160]. This and further deletions of the *H19* coding region [84, 85] have demonstrated that neither the presence of the *H19* gene product, nor transcription of a gene *in cis* to *Igf2* affects its ability to utilise enhancer sequences.

The cases cited above do not directly disprove the existence of competition between *H19* and *Igf2* for shared enhancers, as biallelic expression of both genes in the same cell had not been demonstrated. This definitive refutation of the model came from an experiment in which allele specific transcription of the two genes was resolved at the single cell level [111]. Allele-specific *in-situ* hybridisation for *Igf2* and *H19* mRNA was performed on cells derived from midgestation mouse liver. While monoallelic expression of both genes on different chromosomes was the most frequent observation, a significant proportion of cells (approximately 20%) expressed the two genes from the same chromosome. Loss of imprinting of one member of this gene pair, but not the other, provides additional evidence that the mechanisms by which these genes are imprinted must be separable.

The role of the H19 5' flanking region

As discussed previously, the 5' flanking region of the *H19* gene can confer monoallelic expression upon multicopy transgenes [37, 38, 39]. A 2kb fragment within this

region is a site of developmentally stable differentially methylated DNA [38, 34, 130], and complex chromatin structure [153, 154, 155]. Thorvaldsen et. al. [161] created a deletion of 1.6kb of this region (the DMD) containing the majority of the methylated domain.. Transmission of the DMD deletion through the paternal germline led to the activation of *H19 in cis*, with a corresponding lack of *H19* promoter methylation. There was little effect on *Igf2* expression. Conversely, when the deletion was transmitted maternally, *Igf2* imprinting was lost. The DMD is therefore necessary for the correct regulation of both genes.

To examine the developmental specificity of the DMD region, an allele of *H19* was created in which 7kb of the upstream region, including the DMD was flanked by *loxP* sitessuch that upon the induction of *Cre* recombinase, a conditional deletion could be created [162]. Deletions of the upstream region were created at three different time points: i) in the parental generation, to examine if the region was necessary for germline transmission of the imprint; ii) in early development, to examine if the region was required to maintain the imprint during cell division; and iii) in differentiated cells, to determine if the region is required constantly for correct gene regulation. It was found that, for the *H19* gene, deletion of the paternal 5' flanking region at the first two stages resulted in a relaxation of imprinting, similar to the DMD deletion, but when this region was deleted *in cis* to *H19* in differentiated cells, monoallelic expression was maintained. If the 7kb region was deleted on the maternal chromosome at any of the three developmental stages, *Igf2* imprinting was lost. The experimenters concluded that two activities were present in the region, one that is required to direct the silencing of the paternal *H19* allele, and one to silence *Igf2* on the maternal allele. The *H19* silencer was suggested to act via DNA methylation of the paternal *H19* promoter in early development, possibly to set up the 'closed' chromatin conformation of the paternal allele, which is then stable through subsequent cell divisions. The second activity, which is required *in cis* at all time points to silence the *Igf2* gene, has been proposed to constitute a methylation-dependent chromatin boundary element [27, 163].

An allele-specific silencer of H19

As discussed above, a 1.2kb fragment of the DMD containing DMDB and HS2 (***Figure 2b***) has been shown to function as a silencer in *Drosophila* [158]. Deletion of this 1.2kb fragment in the mouse by gene targeting [164] results in upregulation of the *H19* gene from

the paternal allele when the deletion is *in-cis*. The normally silent paternal allele is not activated in all of the tissues that express *H19*, and instead is limited to the subset of tissues where gene expression is known to be upregulated by the *H19* enhancers. The authors suggest that this 1.2kb region harbours an allele-specific silencer, which interacts specifically with *H19* enhancer elements. This element resides in a differentially methylated domain, and therefore this modification may play some role in the allele-specificity of the silencer. However, as the 1.2kb region can act as a silencer in *Drosophila* where there is no genome methylation [158], other modifications may be responsible. In addition, deletion of this element does not alter the methylation of either the remainder of the DMD region (containing DMDA and HS1, see ***Figure 2b***), or of the promoter-proximal or gene-specific *H19* methylation. While under normal circumstances *H19*-proximal methylation on the paternal allele correlates very well with gene silencing (see above), in this case high levels of *H19* expression can occur in the presence of promoter and gene methylation. Deletion of the 1.2kb region has no obvious effect either *in-cis* or *in-trans* on *Igf2* expression or imprinting, strengthening the proposal by Srivastava et. al. [162] that the mechanisms by which the *Igf2* and *H19* genes are imprinted are separable.

Boundary elements and Igf2 imprinting

Boundary elements have been described in many developmental systems (reviewed in [165]), including the *gypsy* transposon of *D. melanogaster* [166], at the silent mating-type loci in yeast [167], and at the chicken β-globin locus control region [168]. These elements act by insulating genes from local chromatin, thus ensuring discrete transcriptional boundaries between neighbouring genes, and preventing inactivation of adjacent genes by the spread of heterochromatin. A consequence of this boundary activity is that, when placed between an enhancer element and the promoter of a gene, the boundary element will prevent the activation of transcription by the enhancer. Placing the boundary at a location adjacent to, but not between the two elements does not prevent transcription, demonstrating that the activity of these elements is position dependent.

Evidence that a similar activity may be present at the *Igf2/H19* locus came from a study in which mutant mice were created in which the *H19* enhancers had been moved to a location in the intergenic region, downstream of *Igf2* and upstream of the DMD [163]. Mice in which the enhancers were placed in the intergenic region expressed *Igf2* biallelically, and did not express the *H19* gene. The location of the enhancers is therefore critical in the

imprinting mechanism of *Igf2*. It was proposed that the 5' flanking region of *H19* is critical for this position effect, i.e., that on the maternal chromosome, the upstream region provides a boundary between the enhancers and the *Igf2* promoter. On the paternal allele an imprint, most probably CpG methylation, acts to prevent the enhancer blocking activity and hence allow access of the *Igf2* promoter to downstream enhancers. The enhancer blocking activity of the DMD region has been tested in a transgenic system [169]. Reporter gene expression levels were compared between two transgene constructs, one which contained an intact DMD fragment interposed between a reporter gene and the *H19* enhancers, and another construct identical except for a 670bp deletion that removed HS2 (see ***Figure 2b***). Reporter gene expression in neonatal liver in those lines with the DMD deletion was on average 14-fold higher than expression from constructs containing the intact DMD. These transgenes were expressed biallelically however, and were largely unmethylated suggesting that the necessary sequences for imprinting were not present on the transgene. In the light of recent work which shows the HS2 region of the DMD to be involved in *H19* silencing, rather than formation of a boundary element [164], the differences in reporter gene expression between the two transgene constructs could equally be interpreted as due to the presence or absence of a silencer known to be active in endodermal tissues.

The boundary element properties of the DMD region have also been demonstrated in *in vitro* systems. Reporter constructs containing the DMD region between the promoter and enhancer elements have been tested in a variety of stable cell lines [170], [169], and episomal vectors [155]. In these assays, the DMD is able to downregulate reporter gene expression in a position-dependent manner. The hypersensitive regions of the DMD, HS1 and HS2 displayed partial blocking activity, and constructs containing multiple copies of each hypersensitive region increased blocking activity in a copy number dependent fashion [169]. Regions from the human *H19* upstream region were also tested for blocking activity. The human *H19* 5' region contains several copies of two distinct repeat motifs (the *A* and *B* repeats), which are differentially methylated [156]. A fragment of this region containing several copies of the *B* repeat was found to exhibit enhancer blocking activity [169]. The mouse hypersensitive regions, and the human *B* repeat share a 21bp conserved sequence element [156]. This motif contains a 14bp consensus binding site for the protein CTCF [169], [170]. CTCF has been previously shown to exhibit enhancer blocking activity at the chicken β-globin insulator [171]. Sequences from within the human and mouse DMD were shown to

bind CTCF *in vitro*, and mutation of the core 14bp binding site abolished this interaction. Furthermore, methylation of these regions prevents CTCF binding, both in fully methylated and in hemi-methylated DNA [169, 170].

The DMD region clearly contains important regulatory elements for the imprinting of both *Igf2* and *H19*. *Igf2* imprinted appears to be mediated, at least in some tissues, by the action of a methylation-sensitive boundary element at the DMD. This element, when unmethylated, prevents access of downstream enhancers to the *Igf2* promoters, hence the maternal allele of *Igf2* is silent. When methylated, this element cannot bind boundary specific proteins, and is therefore unable to prevent the access of the *Igf2* promoters to enhancer elements and the gene is expressed. *H19* silencing appears to be linked to the formation of heterochromatin at its promoter, thus leading to an epigenetically stable, silent state. The nucleation of this heterochromatin appears to be initiated at the DMD region on the maternal allele, and may arise by the recruitment of silencing complexes by methyl-binding proteins such as MeCP2, and/or by the action of a methylation-independent silencer element.

It is of interest to note that the HS2 region has been implicated in both position-independent gene silencing [164] and enhancer blocking [170, 169, 155] at the *Igf2/H19* locus. A DNA element which under different circumstances can act as an insulator and a silencer has a precedent in the action of the gypsy insulator in *Drosophila* [172]. The gypsy insulator is known to block access to distal enhancers by binding the proteins *suppressor of Hairy wing [su(Hw)]* and *mod(mdg4)*. In wild-type flies, *mod(mdg4)* is essential for the enhancer blocking activity of the insulator DNA. However, reductions in *mod(mdg4)* activity cause the promoter to function as a position-independent silencer. Therefore the accessibility of *trans*- acting factors to boundary element sequences can lead (at least in this case) to a change of function from a boundary element to a silencer. A parallel could be drawn to the HS2 region at the *Igf2/H19* locus, where modulations of *trans*- factor binding are achieved by DNA methylation. In circumstances where there is a high local concentration of bound factors such as CTCF (i.e., the unmethylated maternal allele), HS2 would act as a boundary element; whereas when these factors were at low concentrations, or absent (such as on the methylated paternal allele), HS2 may act as a silencer. Such speculation requires experimental verification, such as in an *in-vitro* system where the levels of HS2 methylation, and CTCF concentration could be varied.

Further imprinting control regions

The DMD may not be responsible for the full silencing of maternal *Igf2* and paternal *H19*, as the boundary element experiments utilising transgenics only test the interactions between the DMD and the *H19* enhancers [169], which do not drive gene expression in all cell types [104]. Furthermore, the deletion of the DMD region only resulted in a partial derepression of the silent *Igf2* and *H19* alleles [161, 162].

When the *H19* coding region is replaced by a *luciferase* reporter gene, in both transgenes [39] and its endogenous location [160], full silencing of this locus on the paternal allele is rarely observed. Sequences within the *H19* coding region may therefore be at least partially responsible for its imprinting status.

At the DMD, both mechanisms of allele-specific expression (the *H19* silencer and the *Igf2* boundary) rely on specific interactions with enhancer elements. These observations raise the question of whether a common feature of imprinting control elements is to limit the access of genes to positive-acting regulatory factors. The action of additional candidate imprinting control elements (such as the differentially methylated regions at *Igf2* (DMRs 0, 1 and 2)) may therefore rely on interactions with undiscovered tissue-specific enhancer elements.

In the case of the *Igf2*-proximal DMRs, several studies have demonstrated that these regions may modulate *Igf2* imprinting in a subset of tissues where expression levels are not affected by the *H19* enhancers. A 30kb transgene in which the *Igf2* coding region had been replaced with a *LacZ* reporter construct, as well as ~15kb of upstream sequence (including DMR1 and DMR0), displayed imprinted expression in one instance (of eight). This effect was not thought to be due to integration of the transgene at an imprinted site [173]. In this case, expression of the *LacZ* reporter was present in the dermomyotomes and their derivatives, following paternal transmission of the transgene. Deletion of the *Igf2* upstream region (including DMR1) in a mouse fibroblast cell line resulted in derepression of the silent alleles of both *Igf2* and *H19* [174]. The *H19* enhancers have been demonstrated to be inactive in mouse fibroblasts [175]. In addition, recent work [176] has demonstrated a direct role for *Igf2* DMR1 in allele-specific silencing of *Igf2*. Deletion of this region *in-cis* results in tissue-specific activation of the maternal *Igf2* allele in mesodermal tissues including heart, kidney and lung. In addition, this deletion results in loss of expression of the placenta-specific *Igf2* transcript from P0. Perhaps as a consequence of this, embryos

carrying the deletion are 29% smaller than their wild-type littermates at birth, but regain weight comparable to their wild-type littermates after birth. The deletion has no effect on *H19* expression or imprinting.

The upstream region of *Igf2* has also been shown to contain an antisense transcript in both mouse [14, 6] and man [177], which is expressed from the paternal allele only, and shares some genomic regions with the *Igf2* gene. While overlapping transcripts have been implicated in the regulation of many imprinted genes (see above), in all cases the antisense transcript is imprinted in the opposite direction to the sense transcript, and therefore not expressed on the same chromosome. The role, if any, of the antisense transcript at the *Igf2* locus in the imprinting mechanism of this gene remains unresolved.

Summary and questions.

The closely linked *Igf2* and *H19* genes are expressed in a wide range of embryonic and extraembryonic tissues. Expression of these genes is downregulated shortly after birth in the majority of tissues, with expression of *Igf2* persisting in the exchange tissues of the brain and human liver, and *H19* expression persisting in skeletal muscle. The expression of these genes in the embryonic endoderm and some mesodermal tissues is controlled from a shared pair of enhancer elements, which lie approximately 7kb downstream of the *H19* gene. Enhancer elements responsible for gene expression in the remaining tissues have yet to be fully characterised.

Igf2 and *H19* are oppositely imprinted, the majority of expression of *Igf2* in all tissues but the choroid plexus and leptomeninges of the brain arises from the allele inherited paternally, and the majority of *H19* gene expression arises from the maternal allele. Both genes are associated with areas of allele-specific methylation. The active, paternal copy of *Igf2* is marked by three regions of methylation, two upstream of the coding region, and one within the body of the gene. The function of this methylation is unclear, but thought not to constitute an epigenetic modification that serves to differentiate the two parental alleles, as these regions are not differentially modified throughout all stages of development. Instead these regions may contain tissue-specific silencer elements, to which the binding of effector proteins is determined by their methylation status. The silent allele of *H19* is methylated at its promoter, and within the coding region of the gene, and the packaging of *H19* into a heterochromatic state correlates well with gene silencing in most circumstances. A region approximately 2kb upstream of the *H19* gene (the DMD) has been shown to be responsible for the silencing of both the maternal allele of *Igf2*, and

the paternal allele of *H19*. This region is methylated on the paternal allele during the majority of development, and also in sperm. This methylation may therefore constitute a primary epigenetic mark, which distinguishes the parental alleles and directs subsequent events that lead to the activation/silencing of the two genes on the appropriate alleles. The continuous presence of the DMD region is required to maintain the imprinting status of *Igf2*. The DMD is also required in early development for the correct methylation of the paternal *H19* promoter, and the silencing of the gene. The two activities within the DMD region can be separated into (at least) two domains. Firstly, the DMD contains an element that can form a boundary between the *Igf2* gene and the *H19* enhancers (and possibly additional enhancers which lie beyond the boundary) to block the interaction of *Igf2* with downstream enhancers, thus leading to its silenced state. When methylated, the DMD loses this activity, allowing expression of *Igf2* on the paternal allele. Secondly, the *H19*-proximal half of the DMD contains an element that can silence the paternal *H19* allele, again by preventing the activity of the *H19* enhancers, though not by a boundary activity, and possibly by a methylation-independent mechanism.

Many questions remain concerning elucidation of the mechanisms of correct expression and imprinting of these genes. The elements demonstrated to play a role in the imprinting of *Igf2* and *H19* to date have a common feature. All elements (the boundary, the *H19* silencer, and the *Igf2* mesodermal silencer) act in a tissue-specific manner due to their interactions with enhancer elements. While the nature of these interactions are not understood at present, it is clear that to understand the imprinting mechanisms at this locus, the elements which drive tissue-specific expression of *Igf2* and *H19* must be characterised and mapped. Therefore, what is the nature and location of additional elements that drive the expression of *Igf2* and *H19* in the extraembryonic and mesodermal tissues, as well as in the exchange tissues of the brain? In tissues with imprinted expression of *Igf2*, such enhancers may lie either beyond the proposed boundary, or further boundaries may exist. How does *Igf2* (and not *H19*) escape imprinting in the choroid plexus and leptomeninges? Again, the location of enhancer elements may be critical if the boundary element model of *Igf2* imprinting is correct. Thirdly, why is genome context so critical for the correct imprinting of these genes, i.e., what additional factors are required to stabilise the epigenetic environment that is proposed to be initiated from the DMD region?

Aims of the project

The CCD

The CCD (for centrally conserved domain, see ***Figure 1***) was initially identified in a study that scanned approximately 130kb of the genomic region containing the *H19* and *Igf2* genes for regions that were sensitive to the nuclease DNaseI. A region 32kb upstream of the *H19* gene was found to display a high level of nuclease sensitivity on both parental alleles. This region was also found to be hypomethylated in comparison to surrounding DNA, and to be conserved in a number of mammalian species [151].

The presence of nuclease sensitive sites (indicative of functional chromatin, see above), as well as its evolutionary conservation, suggests the CCD as a good candidate for an additional regulatory element at this locus. This region could function as a tissue specific enhancer or silencer, or a further imprinting element at this locus. The purpose of this work is to investigate the role of the CCD in the tissue-specific expression and imprinting of the *H19* and *Igf2* genes.

Experimental set-up

Is the CCD able to direct reporter gene expression in the developing embryo?

This study will utilise *luciferase* reporter transgenes created by Ward et. al. [108] in order to examine the expression profile conferred on *Igf2* as a result of *cis* regulatory elements (see ***Figure 3***). In these experiments 96 transgenic lines were created containing a *luciferase* reporter gene driven from the *Igf2* P3 promoter, in combination with additional elements from the *Igf2/H19* locus. The elements used are depicted in relation to their endogenous location in ***Figure 1*** and the reporter constructs derived from these regions are shown in ***Figure 3***. Of particular interest is the role of the CCD in the tissue specific-expression of these genes. The following questions will be addressed:

i) Does the CCD region harbour mesoderm-specific enhancer activity?

ii) Can the CCD reproducibly drive reporter gene expression from P3 in the brain, specifically in the exchange tissues?

iii) Does the CCD interact with the *H19* enhancers, to broaden or limit their range of expression, or modulate quantitative levels of expression?

iv) Is the CCD involved in the imprinting mechanism of *Igf2/H19*, acting either alone or in combination with the *H19* enhancers?

These questions will be answered by extending the analysis of Ward et. al. [108], into a broader range of tissues, to assay for *luciferase* activity in transgenic neonates in a) skeletal muscle and tongue (mesodermally-derived tissues); b) brain (specifically the meninges and choroid plexus); and c) liver and kidney (predominantly endodermally-derived tissues), driven by the ***A***, ***H*** and ***Q*** transgenes (see ***Figure 3***). The expression profile of reporter genes in additional transgenic lines created during the course of this study are also reported.

Use of transgenic mouse models

There are many advantages of using a transgenic system in which to investigate the function of elements thought to be required for correct gene expression. Firstly, transgenic systems provide an *in vivo* system, in which the action of elements in a full nuclear context can be investigated. The expression and imprinting of *Igf2* and *H19* has been primarily carried out using mice as a model system, so experiments in transgenic mice are operating in a well understood genetic system.

Some of the disadvantages of using a transgenic system to study complex regulatory interactions have been highlighted above. The limitations of a transgenic system mainly rest upon the danger of such a system being too reductionistic, i.e. that single regulatory elements alone may act in a different manner than when in combination with their usual neighbours. Additionally, but related is the problem of position effects, i.e. that the behaviour of a transgene may not be due to the *cis* elements present on the experimental construct, but due to elements present at the transgene integration site. This problem can be overcome by comparison of several transgenic lines, that share the same experimental construct, but differ in their sites of integration. However, it should be noted that transgene copy number is potentially an additional confounding variable. All experiments in this study will rely on the use of at least three transgenic lines bearing the same construct in order to control for position effects.

Is the CCD able to direct reporter gene expression in an in vitro *system?*

Transient transfection experiments can be a useful tool in the determination of activity and tissue specificity of putative regulatory elements. The *H19* enhancers were initially characterised by transient transfections of constructs containing the gene region where these elements lie into endodermally derived cell lines [103].

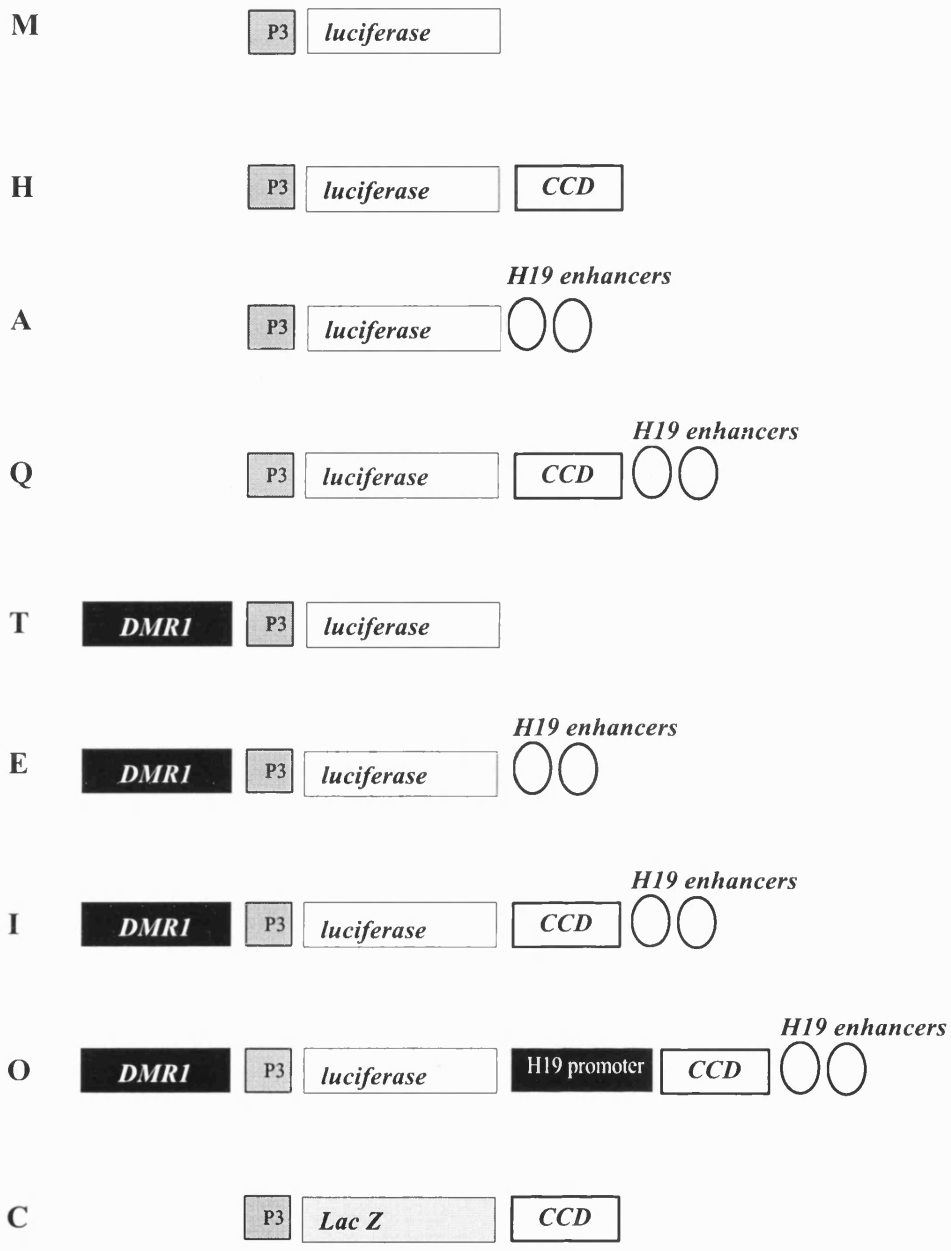


Fig.3 Summary of *Igf2/luciferase* reporter constructs used in this study. All constructs are derived from ***M***, which contains **P3** fused to the *luciferase* reporter gene. Further elements (described in ***Figure 1***) are added to ***M***.

In order these elements are: ***DMR*** (*Igf2* ***DMR1***, approximately 2.8kb *EcoRI-BamHI* fragment); ***P3*** (*Igf2 promoter 3,* nucleotides -162 to +74 with respect to the
transcription start, which includes part of exon 4); ***CCD*** (a nuclease hypersensitive region lying midway between *Igf2* and *H19,* 2kb *EcoRI-EcoRI* fragment); the ***H19 promoter***, (a 2kb *BamHI-NheI* fragment which includes part of *H19* exon 1); the ***H19 enhancers*** (2.7kb SpeI-BglII fragment). The ***C*** is identical to ***H*** except it contains a *LacZ* reporter gene (a 3kb *NotI* fragment) in the place of the *luciferase* reporter gene.

A number of cell lines are used in this study to investigate whether: i) the CCD can drive reporter gene expression in a cell-type specific manner, and ii) the CCD can interact with the *H19* enhancers, to enhance or suppress gene expression.

Can sequence comparisons between the CCD region in a number of mammalian species tell us anything of CCD function?

Evolutionary conservation of the CCD region has been demonstrated by cross hybridisation of genomic DNA from a range of species to a mouse CCD probe [151]. While this study has demonstrated that similar sequences exist in other species, it does not identify what these sequences are. By comparing the CCD sequence from several species, conserved regions within the region can be identified at the nucleotide level, and this information may illustrate the minimal regions required for the function of the CCD.

In summary, the aims of this work are to investigate the function of a conserved, DNaseI hypersensitive region that lies in the intergenic region between the imprinted *Igf2* and *H19* genes. A role for directing both tissue-specific, and allele specific expression will be tested, utilising both *in vivo* and *in vitro* techniques, as well as comparisons of the nucleotide sequence between species.

CHAPTER 2: MATERIALS AND METHODS.

Plasmids

Igf2-luciferase reporter constructs containing candidate regulatory elements from the *Igf2/H19* region (illustrated in ***Figure 1***) were constructed by Dr A. Ward (as described in, [178,108]). The ***M*** construct contains *Igf2* promoter 3 fused to the *luciferase* reporter gene, inserted into the BluescriptSK+ vector. All other constructs were generated by insertion of candidate elements (CCD, DMR1, *H19* enhancers, *H19* promoter) into ***M***. A summary of these constructs is illustrated in ***Figure 3.*** The *Igf2-LacZ* reporter construct containing the CCD region (***C*** lines, ***Figure 3***, **Chapter 1**) was constructed by Dr G. Dell, and is identical to the ***H*** construct except the *LacZ* gene replaces the *luciferase* coding region.
p1*75a:4*, the source of the CCD deletion constructs, is composed of the 2kb *EcoRI* CCD fragment (***Figure 1***) cloned into BluescriptSK+, and was obtained from Dr A. Ward. This plasmid is illustrated in ***Figure 32,*** **Chapter 8.** To create the CCD deletion constructs described in **Chapter 7,** *p175a:4* was digested with *PstI* to produce fragment sizes of approximately 1.1kb, 0.8kb and 3.0kb. The 1.1kb and 0.8kb fragments were then reinserted into the *PstI* cut vector sequence (which contained a residual ~100bp of CCD sequence). This created the truncated and deleted versions of the CCD. The deleted and truncated CCD fragments were then excised from the vector with *SpeI* and *NheI* and these fragments inserted into an *NheI* site immediately downstream of the *luciferase* coding region in the reporter plasmid ***P3MM.*** The deleted (***pCCD4a***) and truncated (***pCCD11a***) CCD reporter constructs are illustrated in ***Figure 26,*** **Chapter 7.**

Transgenic mice

The majority of transgenic mice used in this study are detailed in Ward et. al. [108].

The transgenic mouse lines *Quiche*, *Connie*, and *Columbo* were constructed as follows: Linear construct DNA (1μg/ml) was microinjected into one pronucleus of an *F1* zygote that was the product of a cross between a *C57BL/6J*♂ and a *CBACa*♀, using standard techniques [179]. Injected embryos were transferred into the oviduct of a pseudopregnant *MF1* foster mother by Dr A. Ward or Dr W. R. Bennett.

Lines were established and maintained by breeding the transgenics with *F1 (C57BL/6J x CBACa)* partners. This combination has the advantage of hybrid vigour, as

well as responding well to superovulation, and has been widely used in the production of transgenic mice [179]; but as a consequence the transgene was transmitted through various combinations of the *C57BL/6J* and *CBACa* genomes. The first letter of the name of the transgenic line corresponds to the transgene construct inserted.

Mice with a targeted insertion of a neomycin resistance cassette at the *p53* gene (creating a functional null allele, as described by Clarke et. al.[180]) were obtained from Dr A. R. Clarke, and maintained on an F1 background as described above.

Antibody staining protocols

(Adapted from Harlow and Lane, 1988 [181])

Embryos

Embryos were obtained from crosses of males hemizygous for the transgene against *F1* (*C57BL/6J x CBACa*) non-transgenic females. The gestational age was calculated from the middle of the dark cycle preceding the observation of a vaginal plug. Embryos were dissected from pregnant females on embryonic day 12.5 (e12.5), or day 14.5 (e14.5) of gestation. *Luciferase* assays were performed on the yolk sac of each embryo to determine whether the embryo carried the transgene (see the isolation and identification of samples section, below).

Paraffin embedding

Transgenic embryos and their wild-type littermates were fixed according to four different methods: i) over-night in 4% paraformaldehyde/phosphate buffered saline (PBS); ii) over-night in MEMFA [0.1M MOPS buffer pH 7.4, 2mM EGTA, 2mM $MgCl_2$, 4% formaldehyde]; iii) 5 hr in Zambonis' phosphate buffer [20mM Na_2HPO_4, 70mM NaH_2PO_4, 0.85% paraformaldehyde, 0.15% saturated picric acid]; or iv) 5 hr in 20% ethanol. Samples were dehydrated in an ethanol series (at least 4hr each in 2 x 70%, 80%, 90%, 95% and 2x absolute ethanol) and cleared in Histoclear (National Diagnostics). The samples were then embedded in paraffin wax (Lamb). Saggital sections were cut at 7μm, dewaxed in Histoclear and rehydrated in an ascending ethanol series.

Cryostat embedding

Transgenic embryos and wild-type littermates were rinsed in cold PBS, then embedded in a block of O.C.T. compound (BDH Laboratory Supplies) by immersion in a liquid nitrogen/

isopropanol slurry. Saggital cryosections were cut at 15μm, dried over-night at room temperature, then stored at -20°C.

Antibody staining on paraffin sections

Sections were washed twice in PBS. At this stage, in some experiments samples were permeablised for 30 min in 0.2% Triton X-100. In all treatments the samples were blocked for 1hr in one of a number of blocking agents: i) BLOTTO [5% skimmed dried milk in PBS]; ii) Bovine Serum Albumin (BSA) + Tween [5% BSA (Sigma), 0.02% Tween 20 (Sigma) in PBS]; iii) FCS [5% Fetal Calf Serum (Gibco-BRL) in PBS]; iv) Horse Serum + BSA [10% heat-inactivated Horse Serum (Gibco-BRL), 5% BSA in PBS].

Primary antibody [rabbit polyclonal anti-*luciferase* (Promega, Cortex or Sigma) or rabbit polyclonal anti-neurofilament 200 (Sigma)], in the same blocking agent used in the blocking step, was added for variable amounts of time (usually 1 hr, 2 hr at room temperature or over-night at 4°C). After two 10 min washes in PBST (PBS + 0.1% Tween-20), sections were incubated with the alkaline phosphatase conjugated secondary antibody, anti-rabbit alkaline phosphatase (Vector) for 1 hr or 2 hr. Slides were washed twice for 10 min in PBST. The substrate for alkaline phosphatase was added [Napthol-AS-MX/ Fast Red TR (Sigma)] with 100mM levamisole to block endogenous alkaline phosphatase activity, and allowed to develop for 5 min. In some experiments a peroxidase secondary detection system was used. In these cases, prior to the addition of primary antibody, the slides were incubated for 5min in 3% hydrogen peroxide (Sigma) in tap water, followed by 30min in 0.5% hydrogen peroxide in methanol, then washed 3 x 10min in PBST, in order to quench endogenous peroxidase activity. Following incubation with the primary antibody (as above), the sections were washed for 3 x 10min in PBST. A biotinylated anti-rabbit IgG (Vector) was applied for 1hr, then washed 3 x 10min in PBST. Slides were incubated for 5min at room temperature in ABC kit streptavidin-horseradish peroxidase complex (Vector), prepared according to the manufacturer's instructions. All further steps were carried out according to the manufacturer's instructions for the ABC kit (Vector). In all cases, nuclei were counter-stained with Mayer's haemotoxylin for 5 min. Slides were mounted in Mowiol (Calbiochem) and examined using a Leica DMR microscope.

Negative controls were carried out with the addition of blocker instead of primary antibody in order to discern any non-specific binding by the secondary antibody, and with the addition of blocker in place of the secondary antibody to control for endogenous peroxidase or alkaline phosphatase activity.

Preparation of acetone powder

Mouse acetone powder was prepared by dissection of e12.5 embryos from an F1 x F1 cross. Embryos were homogenised in 0.9% NaCl at 4°C, at ~1g/ml. 8 ml acetone at -20°C was added and the mixture incubated on ice for 30 min with occasional vigorous mixing. A precipitate was collected by centrifugation at 10,000g for 10 min using a Sorvall RC5B centrifuge with SS34 rotor. Supernatant was discarded and the pellet resuspended in fresh acetone, spun down and air-dried at room temperature on clean filter paper. Dried powder was used at a final concentration of 1%, incubated for 1hr with antibody in blocking agent at 4°C

Antigen retrieval

In some experiments an antigen retrieval protocol was used prior to the blocking step in antibody reactions. Two methods were employed, a high temperature antigen unmasking method (obtained from Dr J. Smith, University of Birmingham), and a microwave antigen unmasking method as described by Cuevas et. al.[182].

The high temperature unmasking method was performed as follows: Following dewaxing and rehydration, slides were pressure cooked for 1min in 10mM citrate buffer pH 6.0 (the pressure cooker was filled with sufficient buffer to cover the slides, brought to the boil, then the slides were placed within and the lid sealed, timing began once the pressure indicator rose to visibility). The slides were then allowed to cool while still immersed in buffer, then washed for 3 x 10min in PBST at room temperature.

Antibody staining of cell culture samples

Cells were grown to a confluent monolayer on sterilised coverslips, then washed in PBS several times. The cells were then fixed for 30min in 4% paraformaldehyde/PBS at room temperature, washed in PBS several times, and permeablised by incubation for 30min in 0.1% TritonX-100 (Sigma)/PBS. Blocking was carried out by incubation of the coverslips in blocking buffer [20% Blocking Reagent (Boehringer Mannheim)/0.1% Triton X-100/PBS] for 3hr at room temperature. Primary antibody (rabbit α-prealbumin (transthyretin), DAKO) was added at $^1/_{100}$ in blocking buffer, and the cells incubated overnight at 4°C. The coverslips were washed 3x 10min in PBS in a large volume, then the secondary antibody (fluorescein labelled α-rabbit (goat), Vector) was added at $^1/_{1000}$ in blocking buffer, and the coverslips incubated at room temperature for 3 hr. The secondary

antibody was removed by 3 x 10min washes in PBS. FITC and bright field images were obtained by Dr D. Tosh and X. Shen using a Zeiss 510 Laser Scanning Microscope.

Assay for luciferase specific activity

Isolation and identification of samples

Preliminary experiments

Prior to the antibody staining experiments detailed in **Chapter 3**, *luciferase* specific activities were determined for the offspring of crosses of transgenic fathers (*Alicia*, *Ost* and *Helga* lines) vs. non-transgenic (F1) mothers. Embryos were dissected from the uterus at 12 days following the appearance of a vaginal plug, crudely dissected, and assayed for *luciferase* activity. Samples collected were; liver, brain, tail, tongue and placenta/yolk sac. For identifying transgenic offspring the extraembryonic tissue samples were placed in 100 µl cell lysis buffer [25mM Tris-phosphate pH7.8, 2mM DTT, 2mM 1,2-diaminocyclohexane-N, N, N', N' tetracetic acid, 10% glycerol, 1% TritonX-100]. 5µl was assayed for *luciferase* specific activity as described below.

Tissue-specific and parent of origin-specific expression of transgenes at e14.5 and D1

Samples of whole brain, tongue, muscle (upper hind limb), liver bulk and kidneys, at 1day post-partum (D1), and head, body, yolk-sac and placenta at e14.5 were collected on ice and stored at –20°C in 1.5ml eppendorf tubes. For specific activity assays of each transgenic line, at least 40 samples of each tissue were taken on the day of birth, or at e14.5. Half of these followed paternal transmission of the transgene, the other half from maternal transgene transmission. Both sets were handled simultaneously on ice.

For identifying transgenic offspring, tail tips ~0.5cm long were placed in 100µl cell lysis buffer and 5µl was assayed for *luciferase* specific activity as described in in the following section.

Specific activity assays

Cell lysis buffer (1ml for brain, liver, head and body samples, 0.5ml for tongue, muscle, kidney, placenta and yolk-sac samples) was added to each sample and the tissue disaggregated with the 0.1-5ml dispersing tool of an IKA-Labortechnik Ultra-Turrax T8 homogeniser (about 10s, medium setting). To prevent cross-contamination, the dispersing

tool was rinsed several times in water and wiped dry between each sample. The samples were spun at 13000rpm for 3min and duplicate 5µl amounts were taken from the clear supernatant. These samples were placed in microtitre plates and the enzyme activity measured using 100µl *Luciferase* Assay Reagent (Labtech International) in an Anthos Lucy 1 luminometer (in the case of D1 samples), or a E. G. & G. Berthold Microplate Luminometer LB96V. Light emission was counted for 10s. For each assay, a standard curve was constructed using serial dilutions of purified *luciferase* (Sigma). The relationship of relative light units to amount of *luciferase* was linear in the 100ng to $1x10^{-5}$ ng range, and all samples assayed fell within this range.

Protein assays

Luciferase activity measurements were normalised for soluble protein concentration. 10µl amounts of the homogenised tissue supernatant were diluted 100-fold (tongue, muscle, kidney, head and yolk sac) or 200-fold (brain, liver, body and placenta) with water. From each of these dilutions a 10µl aliquot was mixed with 200µl BioRad dye reagent, and the absorbancy at 590nm measured against dilutions of a BSA standard, as described by Bradford [183]. The specific activity was always expressed as the ng equivalent of the *luciferase* standard per mg soluble protein.

Statistical analysis

For each tissue in each line, the specific activity of the reporter was measured independently for paternal and maternal transmission of the transgene, and then the differences in the means were compared by using a Student's *t*-test [184].

Southern blotting and hybridisation

Genomic DNA preparation

For DNA methylation analyses, genomic DNA was prepared from samples derived from the *luciferase* standard activity experiments (see above). In all other cases, approximately 500mg tissue sample was placed in 560µl tail buffer (50mM Tris.HCl pH 8.0, 100mM EDTA, 100mM NaCl, 1% SDS) containing 285µg/ml Protinase K (Boehringer Mannheim). This was incubated overnight at 55°C.

For all samples, 500µl homogenised or digested tissue (in lysis or tail buffer, respectively) was treated with 0.7µg/ml RNaseA (Sigma) for 1hr at 37°C. NaCl was added

to 1.25M, followed by an equal volume of chloroform/isoamylalcohol 24:1 (vol/vol), and this was gently mixed on an orbital shaking platform for 2hr at room temperature. The solution was spun in a microcentrifuge at 13000rpm for 10min, and the upper, aqueous layer transferred to a fresh tube. An equal volume of propan-2-ol was added to this fraction, and mixed gently to precipitate the DNA. This was spun at 13000rpm for 15min, and the supernatant removed. The pellet was washed for 1hr in 70% ethanol, then the pellet was allowed to dissolve overnight at 4°C in 100µl TE (10mM Tris.HCl pH 8.0, 1mM EDTA).

Southern blotting

~20µg genomic DNA samples were digested overnight with restriction enzymes, run on a 1% agarose gel and transferred onto Hybond-N (Amersham) membranes according to standard procedures [185].

Probe fragments used in these analyses are detailed in the following section. Fragments were gel purified using the Promega gel extraction kit and ~20ng labelled to high specific activity with ^{32}P dCTP (Amersham) with the High Prime labelling kit (Boehringer Mannheim), according to the manufacturer's instructions. The blot was prehybridised for 2 hr in Church's buffer [186] and then approximately$10x10^6$ c.p.m. of denatured probe added. The filters were hybridised for at least 12 hr at 65°C, then washed twice for 15 min and once for 1 hr in wash solution containing 1%SDS, 40mM Na_2HPO_4, 1mM EDTA. All washes were done at 65°C.

For hybridisation to non-identical sequences (as for rat DNA described in **Chapter 8**), hybridisation was carried out at 55°C, and washes were identical to those described above, except they were performed at 55°C.

The filters were then rinsed in 40mM Na_2HPO_4, wrapped in SaranWrap and put down for autoradiography with X-ray film (X-OMAT AR, Kodak) at -80°C with intensifying screens.

Probes used in Southern blots

Probes used in methylation analysis described in **Chapter 5** are illustrated diagrammatically in ***Figure 13*** and ***Figure 14*, Chapter 5**. The P3-*luciferase* probe specific to the *Igf2* P3 promoter, and the 5' region of the *luciferase* gene was prepared by digestion of the ***M***-construct containing plasmid (*p3MM*, see plasmids) with *EcoR1*. The

appropriate 984bp band was purified from a 1% agarose gel using the Qiagen gel extraction kit (Qiagen), according to the manufacturers instructions. An *H19* enhancer probe, spanning the mid- to 3' region of the *H19* enhancer region was cut from an *A*-construct containing plasmid (*p166a:2,* obtained from Dr A. Ward) with *PstI*, and the appropriate 949bp band was gel purified as described above. A CCD probe corresponding to the entirety of the CCD sequence used in the ***H*** construct was excised from *p175a:4* by digestion with *EcoRI*. The appropriate 1910bp fragment was gel purified as above.

Transient expression assays

Cells

The human hepatocarcinoma-derived cell line *Hep3B*,the monkey kidney cell line *Cos7*, and the mouse fibroblast cell line *NIH3T3* were all obtained from Dr. A Ward and cultured at 37°C, 5% CO_2 in Dulbeco's Modified Eagles Medium (DMEM, Gibco-BRL) supplemented with 5% Fetal Calf Serum, 25μg/ml amphotericin, 100IU/ml penicillin, 100 μg/ml streptomycin and 2mM glutamine (all Gibco-BRL).

The murine myoblast cell line *C2* were obtained from Dr J. Smith, and cultured as above, except Fetal Calf Serum concentration was 10%. *C2* cells were induced to differentiate into myotubules as described previously [187], [188]. Briefly, cells were washed three times in sterile PBS, then culture medium (DMEM with antibiotics as above) was added, supplemented with 10% Horse Serum (Gibco-BRL). Myotubules were first visible at approximately 48hrs following the addition of Horse Serum.

The origin of choroid plexus cells is detailed below.

Transient transfections

Transfections of *Hep3B*, *Cos7*, *3T3* and *C2* cells were performed using FuGene reagent (Boehringer-Mannheim) according to the manufacturer's instructions for transient transfections of adherent cells. These cells were plated in 6-well dishes at least 24hr before transfection. In the case of choroid plexus cells, the tissue was prepared as described in Section 7.2, and the cells plated into 6-well dishes at a density of ~$5x10^5$ cells/well, and cultured as described in that section. 24hr prior to the transfection, the medium was replaced with choroid plexus culture medium lacking cytosine arabinoside. All cells (~70% confluent) were transfected with 1μg *luciferase* reporter plasmids *M, H, A* and *Q*, and in some cases the CCD deletion constructs ***pCCD4a*** and ***pCCD11a*** (see plasmids, and

Figure 3, **Chapter 1** and *Figure 27,* **Chapter 7**). The vector *pSVβ-galactosidase* (1μg, Promega) was co-transfected as an internal standard for transfection efficiency. *pBluescriptSK+*(Stratagene) was used as a negative control. The cells were harvested 48hr after transfection and were assayed for *luciferase* and *β-galactosidase* activity (as in [189]). Protein was measured using the Bradford assay with known quantities of bovine serum albumin (BSA) used as a standard [183].

Luciferase activity assays of transfected cells

Cells were rinsed once with PBS then lysed by addition of 100μl Cell Lysis Reagent (Labtech International) for 10min. Cell lysates were then microcentrifuged (13,000rpm, 1 min), and 10μl added in duplicate to microtitre plates. The enzyme activity was measured as described for tissue samples above.

Preparation of primary cultures of mouse choroid plexus

(*A modification of Hoffmann et. al.* [190])

Preparation of tissue culture vessels.

Prior to plating choroid plexus cells, culture vessels were seeded with *NIH3T3* cells, and the cells allowed to grow to a confluent monolayer (for culture conditions see Section 6.1). The cells were then lysed by the addition of 1% TritonX-100 (in large enough quantity to cover the cells) for 30min at room temperature. The culture dishes were then washed 3 times in MilliQ water, and stored at 4°C for up to one month [191].

Choroid plexus primary culture

Approximately 10 choroid plexi were removed from the fourth ventricle of F1 mice, 7-10 days post-partum. The tissue was washed in Mg^{2+}- Ca^{2+}-free PBS, then transferred into 5ml Pronase (4mg/ml in PBS, Calbiochem, as described by [192]) and incubated at room temperature for 5 min. Pronase digestion was stopped by addition of serum-containing culture medium. After centrifugation of the cell suspension at 1000rpm the sedimented cells were resuspended in minimum essential medium (MEM)-Earle's medium with non-essential amino-acids, 10% fetal calf serum, 25μg/ml amphotericin, 100IU/ml penicillin, 100μg/ml streptomycin and 2mM glutamine (all Gibco-BRL). The cells were seeded onto 35mm tissue culture dishes, which had been prepared as described in the previous section. Cells were counted using a haemocytometer, and viability was estimated using exclusion

of 0.01% trypan blue (Sigma). Approximately 1-2x10^6 viable cells were seeded per 35mm culture dish. 48hrs after seeding the culture was washed with PBS to remove erythrocytes and, in order to prevent the growth of cell types other than choroid plexus epithelial cells, the medium was supplemented with 2x10-5M cytosinearabinoside (Sigma, as described by Hoffmann et. al.[190]). Epithelial cells reached confluency after 7-10 days *in vitro* at 37°C with 5% CO_2.

Molecular cloning techniques

Digests

5μg plasmid DNA was digested in a reaction volume of 10-20μl with 10 units restriction enzyme (Promega) over-night, according to the manufacturer's instructions.

Ligations

Ligations were carried out in a volume of 10μl containing 1μg digested plasmid DNA, 10 units T7 DNA ligase (Promega) and ligation buffer. The reaction was carried out at 13°C for 4hr, or over-night. Following incubation, DNA ligase was heat deactivated at 75°C for 10 min. Reactions were stored at -20°C, or used immediately in a transformation reaction.

Transformation of DH5α cells

100μl competent DH5α cells (provided by J. Dutton) were mixed with 2μl ligation reaction. The cells were incubated on ice for 10 min, heat-shocked at 42°C for 90 s, then 1 ml SOC [185] was added, and the mixture incubated at 37°C for 15 min. 100μl of the transformation mix was plated onto LB agar [185] with 50mg/ml ampicillin (Sigma).

Plasmid minipreps

(after Miller and Miller, University of Oregon)

Single colonies were picked from transformation plates and grown overnight in 3ml LB + 50mg/ml ampicillin in a shaking incubator at 37°C. 1.5ml culture was transferred to an eppendorf tube and centrifuged at 13,000rpm for 3 min. The pellet was resuspended in 400 μl resuspension buffer [15% Sucrose, 50mM Tris HCl, 50mM EDTA, pH8.5] and vortexed. 50μl lysozyme (12mg/ml) was added and incubated at room temperature for 5 min. 300μl Lysis buffer [70mM Tris HCl, pH8.0, 70mM EDTA, 50mM EGTA, 0.2% TritonX-100] was added. Tubes were inverted to mix and placed at 70°C for 10min. The

mixture was centrifuged at full speed for 10 min and the lysed cells pelleted were removed. 1µl DEPC (Diethyl Pyrocarbonate, Sigma) was added, and the mixture was placed at 70°C for a further 10min. Tubes were spun for 2min, the supernatant added to a fresh tube and DNA precipitated by tha addition of absolute ethanol [185]. Pellet was resuspended in 55µl TE [10mM Tris HCl pH8.0, 1mM EDTA] + 10µg/ml RNaseA. 5µl was digested with restriction enzymes to confirm plasmid identity.

DNA sequencing

DNA sequencing of plasmids was carried out with, and according to the instructions of the T7 Sequenase Kit (Amersham Life Science) using T7 Sequenase version 2.0 DNA polymerase. 50µl plasmid DNA derived from minipreps was denatured by addition of 5µl 2M NaOH/2mM EDTA, incubation at room temperature for 5 min, then addition of 5µl 2.5M Ammonium acetate + 150µl 100% ethanol, and stored over-night at -20°C. The following day the DNA was precipitated, washed once with 100% ethanol, and allowed to air dry.

The template was labelled with 5µCi ^{35}S and annealed to 1pM of one of the following primers:

-40 M13 (Amersham): 5' GTTTTCCCAGTCACGAC 3'

KS (Perkin Elmer): 5' TCGAGGTCGACGGTATC 3'

T7 (Gibco-BRL): 3' CGGGATATCACTCAGCATAATG 5'

SP6 (Gibco-BRL): 5' ATTTAGGTGACACTATAGAATAC 3'

Primers used for sequencing of the rat CCD were:

CCDP1: 5' CAGRGCYRGGRCAGAGGAAGA 3' (where R is A/G, Y is C/T)

CCDP4: 5' CCAACCTTCCTAACAACCTGC 3' (both Gibco-BRL),

as well as T7 and SP6

The sequencing gel was created with the use of the Sequagel 'Ultra Pure' Sequencing system (National Diagnostics) and run on an IBI 'Baserunner' sequencing rig at 30W for at least 4 hr. The gel was dried and autoradiography performed at room temperature with the X-ray film described in Section 5.2.

In some cases sequencing reactions were carried out by Dr P. Jones (Sequencing Core Facility, University of Bath).

Ribonuclease protection and Northern hybridisation analysis

Isolation of total RNA

Total RNA was isolated according to the manufacturer's instructions using TRI Reagent (acid guanidinium thiocyanate phenol chloroform extraction, Sigma).
50-100mg of tissue (D1: liver, fourth ventricle choroid plexus, brain depleted for choroid plexus, and various embryonic and extra-embryonic tissues as specified in **Chapter 8**) was removed from the animals immediately after sacrifice, placed in 0.5ml TRI Reagent on ice, and homogenised immediately with the 0.1-5ml dispersing tool of an Ultra-Turrax T8 homogeniser (IKA-Labortechnik). Subsequent isolation of RNA from the homogenate was carried out according to the manufacturer's instructions for TRI Reagent. Purified RNA was stored at -80°C in DEPC (diethylpyrocarbonate)-treated H_2O.

Ribonuclease protection assays

All hybridisations were carried out at 48°C.

RNase protection assays were performed on 10μg of RNA using the method of Isaacs et. al [193]. Probes were generated and transcribed as follows: the *luciferase* probe was derived by subcloning a 400bp fragment of *P3MM* (containing the P3 promoter, a 70bp region of *Igf2* exon 3 and the first 160bp of the *luciferase* coding region) into *BlueScript KS+* (Stratagene). This plasmid (*pLuxP*) was digested with *BamHI*, and transcribed with T7 polymerase (Promega), such that the resulting antisense RNA protected 160bp of a *luciferase* transcript in transgenic samples, and 70bp of the *Igf2* mRNA in all samples. The probe for the ubiquitously expressed control *mGAP* was made by digestion of an *mGAP* containing plasmid (obtained from A. Ward, as described in Rathjen et. al. [194]) with *AccI,* and transcribing with T7 polymerase, protecting a 65bp region of the *mGAP* transcript. The CCD probe was generated by subcloning a 313bp *ScaI/EcoRI* fragment of *p175a:4* (containing the CCD open-reading frame) into *BlueScript KS+* (Stratagene), This plasmid (*pCCDP*) was digested with *EcoRI* and transcribed with T3 polymerase (Promega), or digested with *HindIII* and transcribed with T7 polymerase.

Northern blotting and hybridisation

Northern blots were performed according to standard protocols [185] with modifications. RNA samples were separated using standard gel electrophoresis, on 1.2% agarose gels: 4.2g agarose was dissolved in 304ml nuclease free H_2O and cooled to 65°C in a water-

bath. To this was added 35ml 10x MOPS running buffer [10x buffer = 200mM MOPS (3-[N-Morpholino]propanesulphonic acid), (Sigma) pH 7.0; 50mM Na Acetate, 10mM EDTA) and 10.5ml 37% formaldehyde (Sigma)] and the gel was immediately poured. RNA samples were prepared for loading by mixing up to 11.25µl RNA solution with 5µl MOPS buffer, 8.75µl formaldehyde, 25µl formamide (Sigma), in a total volume of 50µl. This was denatured for 15min at 55°C, 5µl 10x Ficoll loading buffer was added [185] and the samples loaded onto the gel. The gel was run in 1x MOPs buffer for 18hr at 50V. The gel was then capillary blotted onto Hybond-N membrane.

Probes used in Northern hybridisation experiments

An *mGAP* [194] probe for quantitation of mRNA levels during Northern analysis was obtained from Dr A. Ward, this probe will hybridise to the 1.5kb *mGAP* transcript. A mouse *Igf2* exon 6 probe was obtained from Dr A. Ward, and is predicted to hybridise to the 2.5kb *Igf2* transcript. To create the *transthyretin* (*TTR*) probe, the plasmid *pTTR-7kbexV3* was obtained from Dr R. Costa (described in Yan et. al.[195]). This plasmid contains 7kb of the genomic region of mouse *TTR*, as well as exons 1 and 2, and intron 1. A probe of 980bp covering exon 1, intron 1 and approximately half of exon 2 was obtained by digestion of this plasmid with *BglII* and *StuI*. This probe would be predicted to hybridise to a 0.7kb *TTR* transcript in liver and choroid plexus tissue, as well as a 1kb choroid plexus-specific *TTR* transcript. For detection of transcripts from the ORF within the CCD, the CCD probe previously utilised for Southern hybridisation was used (see above, Section 5.3). Probe labelling and hybridisation of blots was carried out exactly as described for Southern blotting (see above).

Polymerase chain reaction (PCR) analyses

Amplifying the rat CCD

Primers: *CCDP1*: 5' CAGRGCYRGGRCAGAGGAAGA 3',

CCDP2: 5' CCTTCTGWCCWGCTSCAAGCT 3',

CCDP3: 5' CGTGAGGTCAGYGGKYAGCAT 3',

(where R is A/G, Y is C/T, W is A/T, S is C/G, K is G/T, Gibco BRL).

Initial PCR was performed using Reddy-Mix PCR buffer with 1.5mM $MgCl_2$ according to the manufacturer's instructions (Advanced Biosciences), with primers at 1.5µM.

Subsequent PCRs were carried out using Jeffreys' Buffer [196], with 1 unit of ABI Taq polymerase, and the same primer concentrations as above. PCR reactions were denatured for 5 min at 94°C, then cycled at 94°C for 1 min, 62°C for 1 min, 74°C for 1 min for 36 cycles, followed by a 10 min final extension at 74°C. Completed reactions were run out on 2% agarose gels, stained with ethidium bromide and examined on an ultraviolet light source.

p53 PCR

The primers used to identify *p53* null mice described in Section 1.2 are those described in Malcomson et. al. [197]. PCR was performed using the Reddy-Mix PCR buffer with 1.5mM $MgCl_2$ according to the manufacturer's instructions (Advanced Biosciences), with primers at 1.5μM. PCR cycles were as described above.

CHAPTER 3: IS THE CCD ABLE TO DRIVE REPORTER GENE EXPRESSION *IN VIVO?* QUALITATIVE ASSAY.

Introduction

Background

This study will use the resources of a large panel of transgenic mice, which have been generated by Ward et. al. [108]. These transgenics contain different candidate regulatory elements of the *Igf-2/H19* locus fused to a *luciferase* reporter gene, under the control of *Igf2 promoter 3* (P3, depicted in ***Figure 1*** and ***Figure 3,*** **Chapter 1**).

Expression of *luciferase* from these transgenic constructs reproduced several of the characteristics of expression of the endogenous *Igf2* gene. The transgene activity followed the decline of both *Igf2* and *H19* mRNA that occurs after birth [95]. Assays for *luciferase* in these lines following paternal and maternal transmission demonstrated that the *H19* enhancers regularly imposed imprinted gene expression, as well as confirming that this element is a strong enhancer of P3 in the liver. These experiments have also indicated that the CCD may be an enhancer of P3 in the brain.

Accurate assays of reporter gene activity in whole organs fall short of the ideal, which is measurement of activity in each cell type of every organ. The method of Ward et. al. [108] allowed the detection of minor but significant changes in expression in a large number of samples. Since whole organs were homogenised, it was not possible to distinguish between changes in activity due to a change in the amount of *luciferase* produced by a particular cell type, or a change in the range of cell types expressing *luciferase*.

Dissection of organs to give rise to samples of cell subtypes within that organ are problematic, as one cannot always be sure that such dissections are clean, and in many organs (eg, the kidney), cells of different germ layers are highly intercalated. Far more sensitive methods for identifying gene expression at the cellular level are provided by *in-situ* hybridisation techniques.

The aim of the following experiments was to develop a system in which to detect *luciferase* reporter gene activity *in vivo* by the use of immunocytochemistry.

Experimental strategy

In this study, embryos derived from crosses of transgenic mice can be treated with an antibody raised against the *luciferase* protein, in an attempt to resolve *luciferase* expression. If successful this technique can provide information of both spatial and temporal activities of regions contained in the transgenes. Of most interest is whether any expression of *luciferase* can be resolved in ***H***-line transgenic embryos (the ***H***-construct contains the CCD alone fused to P3-*luciferase*, see ***Figure 3,*** **Chapter 1**).

Commercial sources of the α-luciferase antibody

Immunohistochemistry using anti-*luciferase* antibodies is a technique with few precedents in the literature. Two published studies [198], [199], used this technique to detect transgene expression of *luciferase*, in transgenic, wax embedded tissues of mice. In both cases the antibody used was generated by the workers directly, by raising antibodies in rabbits against firefly *luciferase*, followed by affinity purification. These preparations have not been made generally available. Due to time constraints, it was decided to use a commercial source of α-*luciferase* antibodies, rather than go through the lengthy procedure of creating and purifying this reagent *de-novo*.

Several commercial sources exist for this reagent (Sigma, Cortex and Promega). All three sources raised antibodies in rabbits against purified *luciferase* protein, and the immune sera were then affinity purified on *luciferase*-coated columns.

These anti-*luciferase* antisera were polyclonal preparations, which potentially contain multiple antibodies directed against different epitopes on the *luciferase* protein. The advantage of using a polyclonal antibody preparation is that it is possible for several antibodies to bind to the same protein target, thus generating a strong signal [181]. A disadvantage of using such preparations is that there are relatively high concentrations of antibodies that can potentially bind to epitopes present on other proteins apart from the target (i.e., cross reactivity), generating background signals.

Preliminary experiments

Previous to this work, (C. Mackensie, unpublished data) carried out preliminary studies on a variety of commercially available *luciferase* antibodies, to study the affinity of these reagents to *luciferase* protein. The conclusions of this study were as follows:

i) Using an enzyme-linked immuno-absorbance assay (ELISA), it was found that the Cortex anti-*luciferase* serum, at a concentration of $^{1}/_{500}$ was able to resolve between 10-100pg of purified *luciferase* protein.

ii) Using Western blotting, purified *luciferase* protein could be detected at levels of 2-10ng in a protein homogenate. A *luciferase* signal was detected in the homogenised livers of transgenic mice (independently assayed quantitatively, and shown to contain ~4ng *luciferase*/Western blotting sample). A signal could not be detected in the lung, skin or brain of the same animal (0.86, 0.88, 0.66ng *luciferase*, respectively), suggesting that the limit of detectability for purified and native *luciferase* protein are similar.

It was decided to assay organs from transgenic mice to be used in immunohistochemistry studies, to ascertain if levels of *luciferase* protein are comparable to those that can be detected by Western blotting. It is difficult to translate thresholds of detection by Western blotting into the ability to resolve a signal by immunohistochemistry, as the quantitative levels do not distinguish between a punctate or diffuse distribution of protein. The former would be expected to be readily detected by antibody staining, whereas the latter may be problematic to resolve over any background staining. As a first approximation, though, levels of *luciferase* protein in whole organs at greater than 2ng are assumed to have a good chance of detection by immunohistochemical staining with an anti-*luciferase* antibody.

Results

Dissections of e12.5 embryos and assay for *luciferase* specific activity.

In order to create a developmental profile of *luciferase* expression in transgenic lines, it was first necessary to ascertain whether *luciferase* protein was expressed at detectable levels in the target embryos. The embryos chosen for this quantitative study were from some of the same transgenic lines as those that would be used in subsequent staining experiments. The embryonic stage e12.5 was chosen as it represents the earliest stage at which staining would be assayed. *Luciferase* gene expression from the P3 promoter has previously been shown to rise from midgestation, reaching a maximum before birth [108]. Consequently, it should be noted that e12.5 embryos might be expected to contain the minimum level of gene expression used in this study.

The offspring of crosses of transgenic fathers (*Alicia, Ost* and *Helga* lines) vs. non-transgenic (F1) mothers were dissected from the uterus at e12.5 days. Embryos were

crudely dissected, and the samples were assayed for *luciferase* activity. Samples collected were; liver, brain, tail, tongue (lower jaw) and placenta/yolk sac. The extraembryonic tissue sample was assayed for *luciferase* and used to judge the transgenic status of the embryo.

Representative samples are tabulated below, with standardised quantities of *luciferase* protein given in ng/whole organ. Transgenic samples are shown paired with non-transgenic littermates (***Table 1***).

Expression of *luciferase* protein is in some cases above the threshold of 2ng *luciferase* (*Alicia* and *Ost* liver) therefore would at least be detectable by Western blotting.

Antibody titre

Preliminary experiments using e12.5 and e14.5 embryos derived from crosses of males from four different transgenic lines (*Alicia, Helga, Tilly* and *Ost*) against *F1* females produced a *luciferase* antibody staining pattern in both transgenic embryos and their wild-type littermates. As the non-transgenics cannot produce *luciferase* protein, this pattern must be due to background.

The background staining was always observed in the same subset of tissues regardless of the transgenic line from which the embryos were derived. These tissues were the walls of the heart and the white matter of the presumptive spinal cord.

This staining pattern was also observed in those embryos treated with a positive control primary antibody, anti-neurofilament 200 (which reacts with a structural protein of the central and peripheral nervous system). Treatment with this antibody (at concentrations of $^{1}/_{100}$ and above) also produced a signal from other areas in the embryo, most notably in the dorsal root ganglia. This staining was generally more intense and localised, in contrast to the background, and believed to be a true signal. No staining of this kind was observed in those sections treated with the anti-*luciferase* antibody, even at antibody concentrations of $1/_{50}$.

Two problems therefore needed to be overcome, the background staining needed to be reduced, and the *luciferase* epitopes (those fragments of the protein in the sample that can be recognised by the antibody) required unmasking.

Line	Standardised ng *luciferase*/organ			
	Liver	*Brain*	*Tail*	*Tongue*
Alicia	20.27	0.85	0.56	0.88
Non-transgenic	1.99×10^{-3}	2.27×10^{-3}	1.63×10^{-4}	6.32×10^{-5}
Ost	11.09	0.54	0.37	2.46×10^{-3}
Non-transgenic	1.53×10^{-4}	6.52×10^{-4}	5.48×10^{-5}	1.88×10^{-4}
Helga	1.72×10^{-4}	3.51×10^{-2}	1.21×10^{-4}	4.24×10^{-2}
Non-transgenic	1.72×10^{-4}	1.72×10^{-4}	3.44×10^{-5}	3.44×10^{-5}

Table 1. Quantitative levels of *luciferase* protein in representative embryos from the transgenic lines *Alicia*, *Ost* and *Helga*.

Secondary antibody

In most cases a secondary antibody labelled with alkaline-phosphatase was used, which was detected with a chromogenic substrate. There is some endogenous alkaline phosphatase activity present in mouse tissues (e.g., in liver, bone, kidney, intestine and placenta), but the addition of levamisole is sufficient to block this activity in all tissues except intestine and placenta. In some experiments, a secondary antibody labelled with horseradish peroxidase was used. In these cases endogenous peroxidase activity was quenched with excess peroxide [181]. Controls in which no antibody was applied, and detection reagents were added showed no staining (except the expected intestinal phosphatase) in both cases, therefore background staining was not due to endogenous enzyme activities.

A common cause of background problems is due to the secondary antibody, which is raised against rabbit proteins, cross-reacting with mouse proteins present in the sample. A control was done in all experiments in which primary antibody was substituted with blocking solution, then treated as normal with secondary antibodies. These secondary antibody-only controls showed no signal in all cases, showing that the secondary antibody has no cross reactivity to mouse proteins.

The secondary antibody was titrated to optimise the amount used in the following experiments. The anti-rabbit antibody was able to detect a signal from positive controls (anti-neurofilament 200 at a concentration of $^1/_{50}$) at titres of $^1/_{100}$ and above. The secondary antibody was thus used at concentrations of $^1/_{100}$ and incubated for one hour in all following experiments.

Primary antibody

Those antibodies in the primary antibody preparation which bind to non-*luciferase* epitopes should necessarily be at lower concentration than those binding to *luciferase* epitopes, against which the antibodies were raised. If the background is due to this inappropriate binding it should be possible to minimise it by titration of the primary antibody to a level at which there is significant anti-*luciferase* binding, but very little inappropriate binding.

Sections of *Tilly* e12.5 embryos and wild-type littermates were treated with anti-*luciferase* at concentrations of $^1/_{500}$, $^1/_{300}$, $^1/_{100}$ and $^1/_{50}$ (***Table 2***). The level of staining was judged subjectively as intensity of pink coloration produced by the action of alkaline–phosphatase on its chromogenic substrate.

No signal at all was observed until the anti-*luciferase* concentration reached $^1/_{100}$, and this signal was the background staining pattern. Similar results were obtained using the anti-Neurofilament 200 antibody, though at concentrations of $^1/_{100}$ and above the specific signal described previously was also present.

These results strongly suggest that the background is not due to the presence in the preparation of antibodies binding non-specific epitopes, as the same subset of non-specific antibodies (i.e. those binding heart and spinal cord epitopes) would not be expected to be found in independently derived primary antibody preparations against two different proteins.

Both primary antibodies are used at concentrations of $1/_{100}$ and above in all subsequent experiments, unless stated otherwise.

Blocking reagents

Background problems can be caused by 'non-specific sticking', i.e. due to either the primary or secondary antibody binding to the specimen through interactions which do not involve the antigen-combining site. This can be combated by the use of blocking agents, solutions with high protein concentrations that by pre-incubation with the antibody can pre-absorb this non-specific activity.

A variety of blocking agents were investigated on *Tilly* embryos and wild-type littermates. Sections were incubated for one hour with Bovine Serum Albumin (BSA), BLOTTO, Horse Serum or Fetal Calf Serum (FCS) blocking solutions before antibody treatments (***Table 3***). Both the primary and secondary antibodies were applied to the specimens in the appropriate blocking agent. Anti-*luciferase* was applied to specimens at $^1/_{100}$ or $^1/_{50}$. As high titres of antibody give more background than low (***Table 2***), the blocking power of the reagents tested can be compared by their ability to reduce different levels of background. Non-transgenics were also stained to test whether any specific staining would be revealed if background levels were reduced.

The most efficient blockers were BLOTTO and Horse Serum, which were able to remove all background at antibody titres of $^1/_{100}$, and reduce background at antibody titres of $^1/_{50}$ as compared to BSA and FCS. In subsequent experiments, Horse Serum was used as a blocking agent.

No specific staining was observed in any treatment.

[α-*luciferase*]	Sample	Level of specific staining	Level of non-specific staining
$^1/_{500}$	transgenic	-	-
$^1/_{300}$	transgenic	-	-
$^1/_{100}$	transgenic	-	+
$^1/_{50}$	transgenic	-	++++
-	transgenic	-	-
$^1/_{500}$	non-transgenic	-	-
$^1/_{300}$	non-transgenic	-	-
$^1/_{100}$	non-transgenic	-	+
$^1/_{50}$	non-transgenic	-	+++
-	non-transgenic	-	-

Table 2. Effect of anti-*luciferase* concentration on levels of specific and non-specific (background) staining. Staining levels are indicated by – (no signal) or number of + (+ low / ++++ high).

Treatment	Sample	[α-*luciferase*]	Level of specific staining	Level of non-specific staining
BSA	transgenic	$^1/_{100}$	-	++
	transgenic	$^1/_{50}$	-	++++
	non-transgenic	$^1/_{100}$	-	+
	non-transgenic	$^1/_{50}$	-	++++
BLOTTO	transgenic	$^1/_{100}$	-	-
	transgenic	$^1/_{50}$	-	+++
	non-transgenic	$^1/_{100}$	-	-
	non-transgenic	$^1/_{50}$	-	++
Horse Serum	transgenic	$^1/_{100}$	-	+
	transgenic	$^1/_{50}$	-	++
	non-transgenic	$^1/_{100}$	-	-
	non-transgenic	$^1/_{50}$	-	+
FCS	transgenic	$^1/_{100}$	-	++
	transgenic	$^1/_{50}$	-	+++
	non-transgenic	$^1/_{100}$	-	-
	non-transgenic	$^1/_{50}$	-	+++

Table 3. Effect of different blocking solutions on the level of specific and non-specific staining with the anti-*luciferase* antibody. Levels of staining are indicated as in ***Table 2.***

Acetone powder

Another method of reducing non-specific sticking is to pre-incubate the antibody with an acetone powder made using the control tissue. This method can also be used to remove antibodies in the preparation that cross-react with non-specific epitopes.

A dehydrated preparation was made from non-transgenic e12.5 embryos with the same genetic background as transgenics, i.e. an *F1* x *F1* cross. Both anti-*luciferase* and anti-neurofilament 200 antibodies (at a concentration of $^{1}/_{50}$) were pre-incubated with this preparation prior to application to the sections. Any protein present in non-transgenic embryos which binds the anti-*luciferase* antibody should be removed by this treatment, leaving only those antibodies which react specifically to epitopes present in the transgenic but not non-transgenic preparations.

The anti-neurofilament 200 treatment acts as a control for this procedure; neurofilament 200 antigen is present in non-transgenics, therefore no anti-neurofilament staining should be observed if the acetone powder treatment is successful.

Results are presented in ***Table 4***. Treatment with acetone powder was able to significantly reduce background signals, and to remove anti-neurofilament 200 specific staining.

Again, no specific anti-*luciferase* patterns were observed in any treatments.

Fixatives

Background signals can be significantly reduced with the use of appropriate blocking agents and treatment with acetone powder. In no treatment so far discussed had any clear specific staining been uncovered with anti-*luciferase* antibodies following the reduction in background staining. The method of fixation employed in all treatments had been that of paraformaldehyde fixation. It was decided to vary the method of fixation to try to unmask any epitopes that might have been hidden by this procedure.

A perfect fixation technique would immobilise the antigen while retaining authentic cellular and subcellular architecture, and permitting unhindered access of the antibody to all cells and subcellular compartments. No fixation technique reaches this ideal, by cross-linking proteins present in the sample many epitopes will be masked or altered by fixatives. Three different fixation techniques which were compatible with wax embedding, in addition to paraformaldehyde fixation, were employed on sections from *Tilly* e12.5 embryos and their wild-type littermates. These fixatives are MEMFA, 20% ethanol and

Zambonis' Phosphate Buffer (detailed in the **Materials and Methods** section), and the results are shown in ***Table 5.***

Very little difference was observed between different fixation techniques. None showed any specific staining with the anti-*luciferase* antibody, while all showed some degree of specific staining with the positive control (anti-neurofilament 200). The most obvious differences were in section quality, the level at which morphological architecture was preserved after treatments. Those embryos fixed in 20% ethanol showing very poor quality, while those fixed in Zambonis Phosphate buffer showed superior quality.

Cryostat fixation

All fixation techniques employing paraffin embedding have failed to unmask any specific anti-*luciferase* staining. Consequently cryostat embedding, a technique with less harsh treatment of tissue was examined. Using this method, no fixation is employed and instead the embryo is snap frozen in the polymeric matrix O.C.T. then sectioned. Any proteins present are assumed to adopt a more physiological conformation as compared to a technique that employs chemical cross-linking [181].

Anti-*luciferase* specific staining was not uncovered by this procedure (in embryos of crosses *Ann* x *F1* and *Holly* x *F1*). However, positive controls have shown anti-neurofilament staining at a level significantly above background, at lower antibody concentrations, suggesting that cryostat embedding techniques are more sensitive than those previously employed, at least for this antibody.

Antigen retrieval

Cuevas et.al. ([182]) made a study of the effect of microwave irradiation on the staining of a range of commonly used primary antibodies in formalin-fixed, wax embedded material. They found that microwave irradiation permitted successful immunostaining with 20 antibodies that otherwise stained only frozen tissues. The staining characteristics of 21 antibodies that were already known to stain formalin fixed, wax embedded material were improved. Another 39 antibodies did not show enhanced staining with this technique.

Treatment	Sample	Antibody	Level of specific staining	Level of non-specific staining
-	transgenic	α-*luciferase*	-	++++
+acetone powder	transgenic	α-*luciferase*	-	+
-	transgenic	α-neurofilament	++	+++
+acetone powder	transgenic	α-neurofilament	-	+
-	non-transgenic	α-*luciferase*	-	+++
+acetone powder	non-transgenic	α-*luciferase*	-	+
-	non-transgenic	α-neurofilament	++	+++
+acetone powder	non-transgenic	α-neurofilament	-	-

Table 4. Effect of acetone powder on the specific and non-specific staining of embryos treated with anti-*luciferase* and anti-neurofilament 200 antibodies. Level of staining is indicated as in ***Table2.***

Treatment	Sample	Antibody	Level of specific staining	Section Quality
Paraformal -dehyde	transgenic	α-*luciferase*	-	good
	non-transgenic	α-*luciferase*	-	
	transgenic	α-neurofilament	++	
	non-transgenic	α-neurofilament	++	
MEMFA	transgenic	α-*luciferase*	-	good
	non-transgenic	α-*luciferase*	-	
	transgenic	α-neurofilament	+++	
	non-transgenic	α-neurofilament	+++	
20% ethanol	transgenic	α-*luciferase*	-	poor
	non-transgenic	α-*luciferase*	-	
	transgenic	α-neurofilament	-	
	non-transgenic	α-neurofilament	+	
Zambonis Phosphate Buffer	transgenic	α-*luciferase*	-	very good
	non-transgenic	α-*luciferase*	-	
	transgenic	α-neurofilament	++	
	non-transgenic	α-neurofilament	+++	

Table 5. The effects of four different fixation techniques on the level of specific staining with anti-*luciferase* and anti-neurofilament antibodies, and on section quality. Levels of staining are scored as in ***Table 2.***

The mechanism of antigen retrieval achieved by microwave irradiation remains obscure, though it has been suggested that microwaves disrupt the cross-linking of proteins.

A further method of antigen retrieval is by a high temperature technique, in which specimens are heated at high pressure. This method has also brought favourable results with a number of antibodies that do not normally stain fixed, wax embedded tissues (J. Smith, pers. comm.).

Both methods of antigen retrieval were attempted on a number of wax embedded sections fixed in Zambonis' phosphate buffer, and stained with α-*luciferase* and α-neurofilament, at antibody concentrations of $^{1}/_{500}$, $^{1}/_{200}$, $^{1}/_{100}$, $^{1}/_{50}$.

Sections treated with the α-neurofilament antibody had a stronger specific staining pattern after antigen retrieval, at antibody concentrations of $^{1}/_{100}$ or higher. Background staining was unchanged.

α-*luciferase* staining was not revealed by either antigen retrieval technique. Microwave irradiation substantially increased the level of background staining, especially when the antibody concentration was high.

Different brands of antibody

All of the techniques described above were carried out using the α-*luciferase* antiserum from Cortex, for which the initial ELISA and Western blot characterisations were done.

Subsequently, several of these trials were repeated with antibodies from Promega and Sigma. These antibodies have been used in combination with paraformaldehyde and Zambonis'-fixed wax embedded sections, with the best of the blocking reagents (Horse-Serum). Cryostat embedded tissue was stained with Sigma α-*luciferase*, but not the Promega reagent, as by this point it had been withdrawn from sale. In all cases the background staining at high antibody concentrations was identical to that observed with the Cortex antibody, and no specific staining was ever observed.

Conclusions

Immunohistochemistry is a powerful technique for discovering information about the spatial and temporal expression of protein encoding genes.

It has been found that transgenic mice with a variety of regulatory regions of the *Igf2/H19* locus fused *in-cis* to a *luciferase* reporter construct can demonstrate readily detectable levels of gene expression when measured by a photometric enzyme assay (Ward et.al. [108], and this study). Knowledge of specific domains of expression driven from

luciferase transgenes was limited by the ability to dissect particular tissues in sufficient quantity and purity for the photometric assay.

Immunohistochemical staining of transgenic samples for *luciferase* protein seemed to be the ideal method by which to discover the fine distribution patterns of gene expression from transgene constructs. In this investigation it was attempted to discover the sub-organ distribution of *luciferase* protein *in-vivo.* Unfortunately all attempts to resolve reporter gene expression by antibody staining have failed.

Embryos at e12.5 and e14.5, from transgenic lines known to express abundant levels of *luciferase* protein in a subset of tissues were treated with α-*luciferase* antibodies from a variety of sources, and with an α-neurofilament-200 positive control. While the positive control antibody gave a consistent, neuron-specific pattern of staining under a number of conditions, staining with α-*luciferase* exhibited a general, non-specific background staining pattern, which was identical in both transgenic embryos and their wild-type littermates. Variation in conditions, by use of several fixation, blocking and embedding protocols was able to reduce this background, but never able to reveal a consistent pattern of *luciferase* staining, even when antigen unmasking protocols were employed.

Previous workers have reported α-*luciferase* staining of transgenic, wax embedded, paraformaldehyde/formalin-fixed transgenic tissue[198, 199]. Lee et. al. [200] created a rat cardiac myosin light chain (MLC) 2 promoter-*luciferase* fusion gene, which was introduced by microinjection into 1-cell embryos to create transgenic mice. *Luciferase* was detected by immunostaining in these animals, exclusively in the ventricular compartment of the heart. Another group created transgenic mice using an construct consisting of *luciferase* fused to the murine preproendothelin-1 promoter, and reported *luciferase* staining in the endothelial cells of both large and small arteries, as well as low levels of staining in veins and capillaries [198].

Both groups used antisera preparations that were not from commercial sources [198, 199, 200], the groups having raised the antibodies against *luciferase* in rodents, and purified them using affinity chromatography. It is possible that here lies the difference between the two published studies, and the work presented above. In the literature accompanying the commercial antibodies used in this study, there is no mention of their use on fixed tissues. A possible reason for the failure of this technique lies in the use of commercially available antibodies, preparations which are not sufficiently sensitive to resolve *luciferase* epitopes

in sectioned tissues. A further explanation is that the fixation techniques employed led to the modification of *luciferase* epitopes beyond the recognition of these antibody preparations. This explanation seems less likely as even unfixed cryosections did not give a specific *luciferase* signal.

The two studies cited above also detailed quantitative measurements of *luciferase* enzyme activity, following expression of the reporter constructs in target tissues. Unfortunately in both cases enzyme activity was expressed as relative light units generated/mg protein. As luminometer equipment varies it is not possible to compare these non-standardised measurements with the *luciferase* specific activities measured from *Igf2/H19* reporter constructs. Despite this, it could be suggested that the level of protein produced from preproendothelin-1 promoter-*luciferase* constructs and MLC-2 promoter-*luciferase* constructs is much higher that that produced by any of the transgenic lines assayed here. In this case it may be that *luciferase* protein levels from *Igf2/H19* transgenic lines in this study are too low to be detected by an immunohistochemical method.

Further work

Information about the elements directing the tissue-specific distribution of *Igf2* and *H19* gene expression at single cell resolution will be invaluable in the investigation of the mechanism of imprinting of these genes. The purpose of this study was to investigate whether several elements, which lie *in-cis* to the *Igf2* and *H19* genes, could confer tissue-specific expression upon a *luciferase* reporter gene, by the use of a immunohistochemical strategy. Unfortunately, the detection of *luciferase* protein was unsuccessful in the hands of this investigator, despite extensive optimisation of the protocol. The failure of this technique could be due to several possible causes, and these causes must dictate how to proceed with the investigation of the role of *cis*-acting elements (particularly the CCD) in the regulation of *Igf2* and *H19*. One possible cause is that the commercially obtained *luciferase* antisera was unable to detect *luciferase* protein in tissue sections. This failure could either be due to poor sensitivity, or due to the fact that these antisera had been raised against purified protein, which may display different epitopes than *luciferase* protein produced by mammalian cells. In the transgenic lines discussed in this study, levels of *luciferase* gene expression could be very low, or the gene products too unstable to allow detection by an immunohistochemical method.

These problems could be overcome by the use of one of a variety of techniques. Firstly, an anti-*luciferase* preparation could be generated *de-novo*. Previous investigations

of *luciferase* gene expression *in-situ* have been accomplished by the use of antisera generated by the investigators themselves [199, 200, 198]. However, commitment to generating a new anti-*luciferase* antibody would have to be considered carefully, for considerable time and resources would be involved in creating such a reagent. Furthermore, the use of such a reagent has no guarantee of success, especially if the cause of failure lies elsewhere. Secondly, the search for *luciferase* protein could be abandoned in favour of an assay for *luciferase* mRNA *in-situ*. The use of this technique would dispense with the need for unreliable reagents, and overcome the problem of mammalian cell-specific *luciferase* epitopes. However, *in-situ* hybridisation for mRNA in mouse tissues is another technique that may require extensive optimisation, complicated by the possibility that if the level of *luciferase* protein is limiting, this could be reflected in the abundance of the mRNA. Protein levels are not linear with respect to mRNA levels, as one mRNA molecule is thought to be processed by many ribosomes. A small amount of protein in these samples may therefore reflect an even smaller amount of mRNA, which could be below the threshold of detection of this technique. A third strategy is the creation of new transgenic lines bearing a different reporter gene. *β-galactosidase* has been used successfully as a reporter gene in several studies of gene activity conferred by elements at the *Igf2/H19* locus. In particular this reporter was used to demonstrate the tissue specificity conferred by the *H19* enhancers [106, 109]. *β-galactosidase* is a very stable protein, and its enzyme activity can be measured both quantitatively (by a photometric assay which measures enzyme hydrolysis of the substrate ONGP (*o*-nitrophenyl β-D-galactopyranoside)), or histochemically (by cleavage of the substrate X-Gal (5-bromo-4-chloro-3-inodyl β-D-galactosidase, which creates an indigo blue derivative). A disadvantage of the use of this reporter is that some mammalian cells possess endogenous *β-galactosidase* activity, but this variant is lysosomal and active only at low pH. Endogenous *β-galactosidase* activity can therefore be blocked by performing the assays at a pH greater than 7.5 (reviewed in [201]).

In order to circumvent the problems of visualising *luciferase* gene activity, it was decided to embark upon the creation of new transgenic lines bearing the *β-galactosidase* gene as a reporter, in place of *luciferase*. To date, over 500 embryos have been microinjected with the ***C*** construct (which is identical to the ***H*** construct, excepting the replacement of *luciferase* with *β-galactosidase*), and transplanted into the uterus of pseudopregnant foster mothers. Of these embryos, more than 10% (61) were recovered at D1, and two individuals

have been recovered with successful integration of the transgene (the transgenics *Connie* and *Columbo*, data not shown). Unfortunately time constraints have prevented any further analysis of ***C*** transgenics, so it is not yet known whether these individuals will transmit the transgene to their offspring, or the resulting pattern of *β-galactosidase* activity generated in the presence of the CCD. Ultimately several ***C***-construct transgenic lines will be needed to assess the role of the CCD in directing gene expression to specific tissues.

CHAPTER 4: IS THE CCD ABLE TO DRIVE REPORTER GENE EXPRESSION *IN VIVO?* QUANTITATIVE ASSAY.

Introduction

As was previously discussed in the **Introduction**, the only enhancers at the *Igf2/H19* locus fully characterised to date are those that lie approximately 10kb downstream of the *H19* gene (the *H19* enhancers). These enhancers certainly confer high levels of expression to both genes in endodermally-derived tissues (namely liver and gut epithelium), and may play a role, perhaps in concert with other elements, in driving tissue-specific expression in the yolk-sac and the exchange tissues of the brain.

The following experiments seek to determine whether the CCD has the ability to drive tissue-specific expression of *Igf2* and *H19*, either alone, or in combination with the *H19* enhancers.

Transgenic mice with *luciferase* reporter constructs (detailed in ***Figure 3*, Chapter 1**) were used in an *in vivo* assay to ascertain the quantitative level of gene expression from *Igf-2* promoter 3 (P3) when other *Igf-2* control elements were present. Previous experiments have shown that a construct containing P3-*luciferase* alone (the ***M***-construct, see ***Figure 3***) cannot drive reporter gene expression above the background levels of this assay [108]. Elements of interest in this study are the *H19* enhancers and the CCD. To this effect, expression levels were measured in tissues derived from transgenic mice bearing the ***A*** construct (P3-*luciferase-H19* enhancers), in the lines *Alicia, Archy, Ayah*; the ***H*** construct (P3-*luciferase*-CCD), in *Harold, Holly, Hamish* lines; and the ***Q*** construct (P3-*luciferase*-CCD-*H19* enhancers), *Quasar, Quark, Quiche* lines. Expression levels were examined in three transgenic lines representing each construct in an attempt to distinguish real effects on P3 due to the element of interest, from position effects caused by transgene integration. Any effects observed across the majority of lines bearing a particular element will be regarded as real.

All tissues were collected from one-day post-partum (D1) offspring of crosses of females or males hemizygous for the transgene vs. F1 (the background strain, see **Materials and Methods**) partners. Offspring therefore inherited the transgene from one parent only. Previous experiments [108] have shown that transgene expression levels from

the *P3* promoter are usually maximal from embryonic day 17.5 to 3 days post-partum, after which expression levels decline sharply. At least 20 individuals were collected and assayed from each parental cross to enable comparison of parent-of-origin specific expression within each line by a student's *t* test (see next chapter).

Results

Tissue-specific expression at D1

The first question posed by this experiment was; what tissue-specificity is conferred by the *H19* enhancers and the CCD when linked to the reporter gene-P3?

Previous studies [108] have shown that the *H19* enhancers are able to drive high levels of reporter gene expression from P3 in liver, and moderate levels in heart and kidney. The CCD has been shown to drive expression from P3 in the brain, particularly in the exchange tissues.

The purpose of this study is to extend this previous analysis into a wider range of tissues. The CCD is a candidate for the yet undescribed mesoderm specific enhancers. Reporter gene expression was therefore examined in muscle and tongue, as well as brain, liver and kidney.

A second question within this category is; do the two elements interact to modulate expression levels or create new patterns of expression? This question is examined using transgenic mice bearing the ***Q*** construct, which contains both the *H19* enhancers and the CCD, and has not previously been analysed for tissue-specific reporter gene expression. The *Quiche* transgenic line was created during the course of this study in order to allow examination of three independent ***Q*** lines.

Figure 4 shows tissue-specific reporter gene activity across the lines. It should be noted that expression levels from *Quark* and *Hamish* are extremely low in all tissues, and may represent position effects. Expression following male transmission and female transmission of the transgene are shown separately for each transgenic line, as in some cases the means are significantly different (see **Chapter 5**).

Mesoderm-specific expression

In tongue and skeletal muscle samples the reporter gene is expressed at low levels in lines containing both elements (see ***Figure 4a*** and ***4b***). There appears to be little difference in

levels of reporter gene activity driven by the *H19* enhancers, the CCD or with both elements together.

Endoderm-specific expression

In liver and kidney samples, significant levels of reporter gene expression are only found in those lines that contain the *H19* enhancers (*Ayah, Archy, Alicia, Quark, Quasar, Quiche;* ***Figures 4c*** and ***4d).*** Levels of *luciferase* expression in the liver are extremely high in comparison to expression levels from other tissues.

Brain-specific expression

In brain samples significant levels of reporter gene expression are only seen in those lines containing the CCD (*Quasar, Quiche, Harold, Holly;* ***Figure 4e),*** though a low level of gene expression can be detected in all three ***A*** lines.

To summarise; the *H19* enhancers are able to drive expression of *luciferase* from P3 in tongue, muscle, kidney, and to very high levels in liver. The CCD is able to drive gene expression from P3 in tongue, muscle, and with the greatest upregulation in the brain. The large variation in expression levels between the lines bearing each construct (particularly the ***Q*** lines), has prevented the observation of any interaction between the two elements to alter levels of expression. However, no new patterns of expression were observed when both elements were present in the same construct.

Is the brain expression confined to the exchange tissues?

It was of interest to discover whether the brain-specific expression of *luciferase* was reflective of endogenous *Igf2* expression, or merely the product of ectopic gene expression. *Igf2* expression in the brain is largely confined to the exchange tissues (i.e., the choroid plexus and leptomeninges, see the **Introduction**). *Luciferase* gene expression from transgenic lines would therefore be expected to be concentrated in these tissues, and depleted in the remainder of brain tissues. To test this assertion, brains and livers were removed from *Holly, Alicia* and *Quasar* neonates. The brain sample was further dissected into a sample containing the majority of the exchange tissues (labelled ***CP***), and the remainder of the brain tissue (labelled ***brain***). RNA was extracted from these samples, and gene expression was examined using the ribonuclease protection technique. Sample RNA

was hybridised with a *luciferase* probe (to measure transgene expression) and an mGAP probe to control for loading. The results of this analysis are shown in ***Figure 5***.

Figure 5a shows *luciferase* expression in *Alicia* neonates. Reporter gene expression is very abundant in ***liver*** samples, as might be expected from the analysis of *luciferase* enzyme activity described above. Furthermore, *luciferase* mRNA can be detected in exchange tissue samples, but not in the remainder of the brain. *Alicia* neonates are known to express the reporter gene in the brain (see ***Figure 4e***). The reporter gene expression pattern closely resembles that of the endogenous *Igf2* gene, suggesting that the *H19* enhancers are able to upregulate gene expression in the exchange tissues of the brain. ***Figure 5b*** shows an identical pattern of gene expression for *Quasar* samples as that described for *Alicia* samples. ***Figure 5c*** demonstrates that the gene activity observed in the brains of ***H***-construct bearing transgenic lines is a consequence (for *Holly* at least) of gene expression in the exchange tissues. CCD-bearing lines therefore demonstrate a pattern of reporter gene expression that resembles endogenous *Igf2* expression. It appears that both the CCD and *H19* enhancers are able to drive gene expression in the exchange tissues of the brain, though the two elements may be responsible for gene expression in different cell types (e.g., meningial cells vs. choroid plexus epithelium). Unfortunately, a quantitative comparison could not be made of expression levels between the three transgenic lines, as reporter gene expression levels varied a great deal between different individuals of the same line (e.g., the two *Alicia* samples in ***Figure 5a***).

Figure 4. Tissue-specific expression across the transgenic lines at D1

Figure 4a. Matched expression levels of mean reporter gene activity in tongue samples across the lines, following male (M) and female (F) transmission of the transgene. Reporter gene expression is seen in lines that are representative of all three reporter constructs (***A, Q*** and ***H***). Error bars show the standard error of the mean.

Figure 4b. Matched expression levels of mean reporter gene activity in skeletal muscle samples across the lines, following male (M) and female (F) transmission of the transgene. Reporter gene expression is seen in lines that are representative of all three reporter constructs (***A, Q*** and ***H***). Error bars show the standard error of the mean.

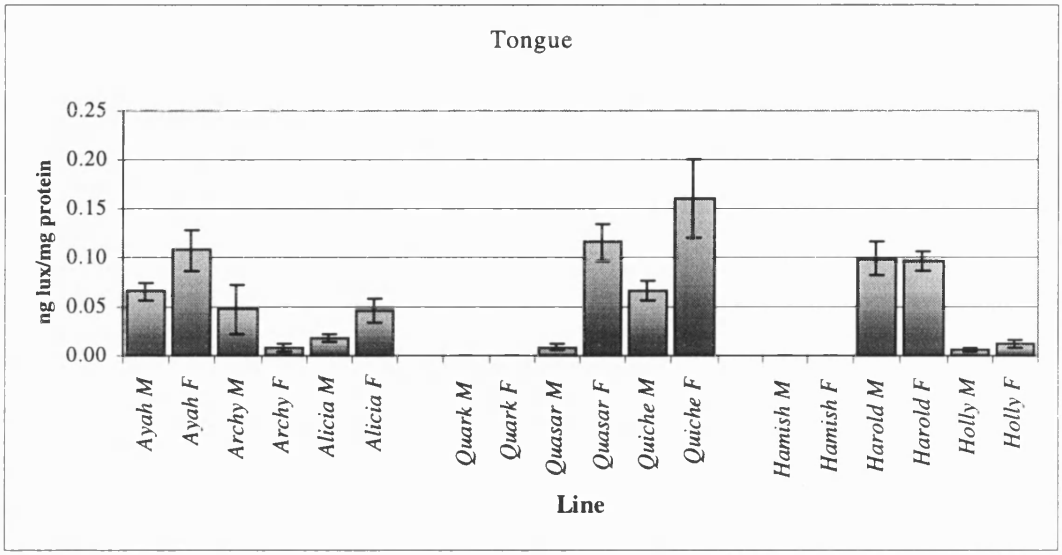


Figure 4a

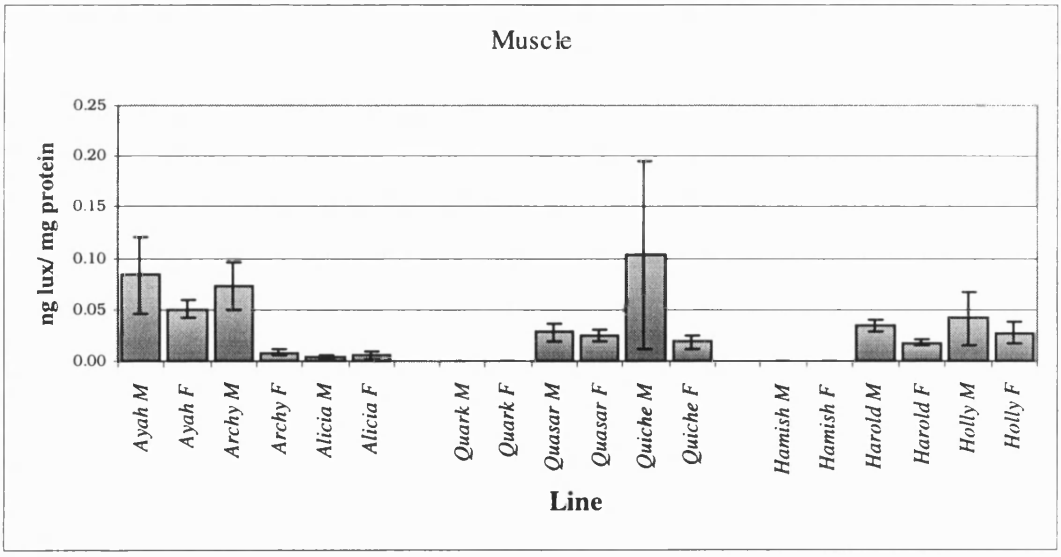


Figure 4b.

Figure 4 (cont.) Tissue-specific expression across the transgenic lines at D1.

Figure 4c. Matched expression levels of mean reporter gene activity in liver samples across the lines, following male (M) and female (F) transmission of the transgene. Reporter gene expression is only seen in those lines that contain the *H19* enhancers, i.e., *Ayah, Archy, Alicia, Quark, Quasar, Quiche.* Error bars show the standard error of the mean.

Figure 4d. Matched expression levels of mean reporter gene activity in kidney samples across the lines, following male (M) and female (F) transmission of the transgene. Reporter gene expression is only seen in those lines that contain the *H19* enhancers, i.e., *Ayah, Archy, Alicia, Quark, Quasar, Quiche.* Error bars show the standard error of the mean.

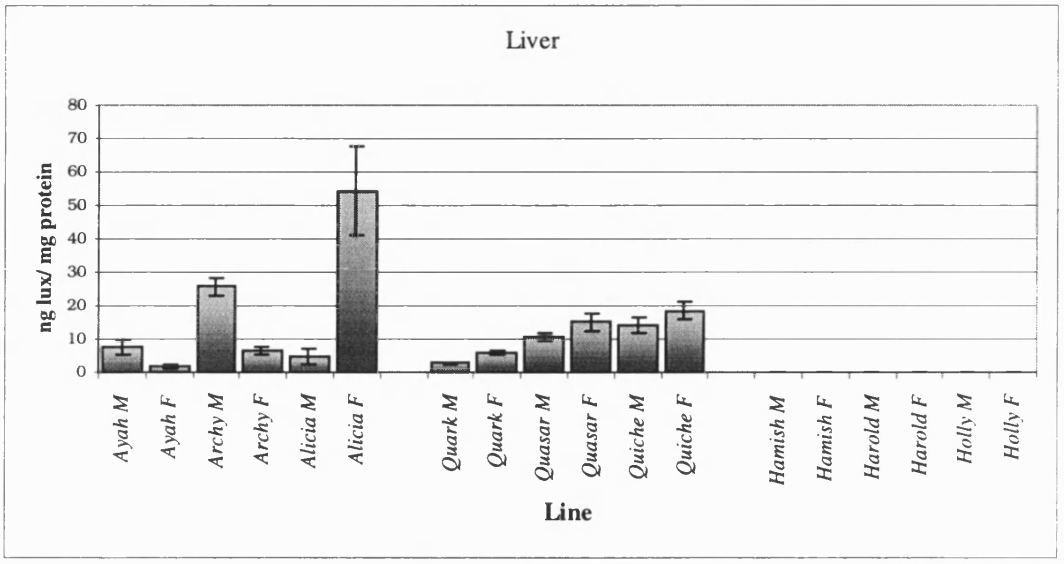


Figure 4c

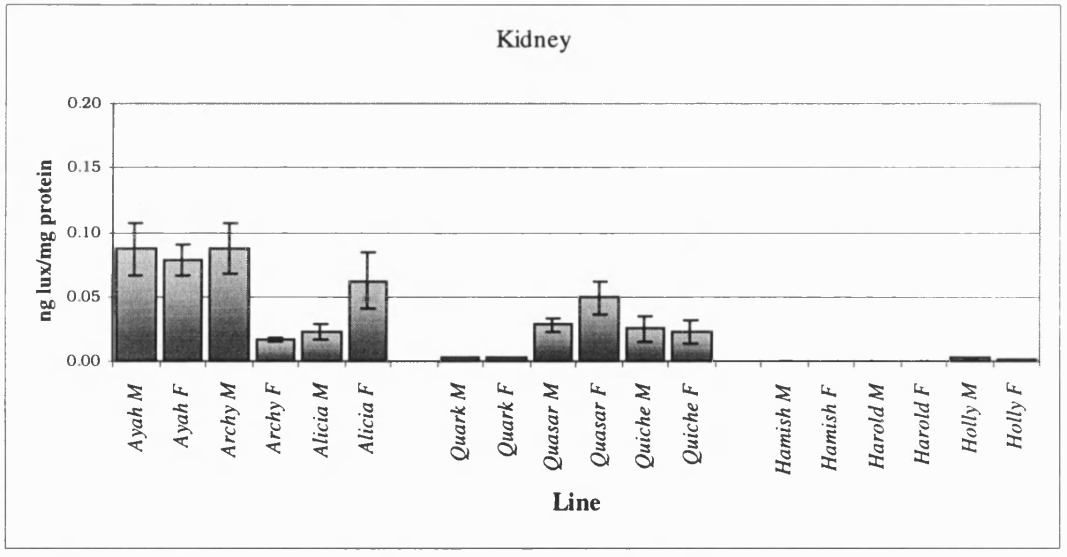


Figure 4d

Figure 4 (cont.). Tissue-specific expression across transgenic lines at D1.

Figure 4e. Matched expression levels of mean reporter gene activity in brain samples across the lines, following male (M) and female (F) transmission of the transgene. The highest levels of reporter gene expression are seen in those lines that contain the CCD, i.e. *Quasar, Quiche, Harold and Holly.* Error bars show the standard error of the mean.

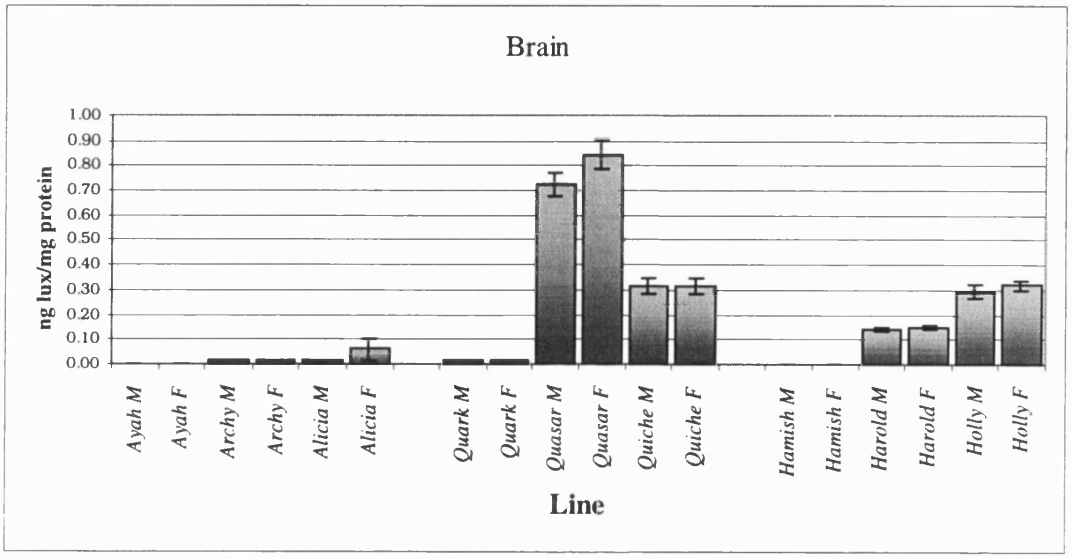

Brain
ng lux/mg protein
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Ayah M
Ayah F
Archy M
Archy F
Alicia M
Alicia F
Quark M
Quark F
Quasar M
Quasar F
Quiche M
Quiche F
Hamish M
Hamish F
Harold M
Harold F
Holly M
Holly F
Line

Figure 4e.

Figure 5. Transgene expression in the brain is confined to the exchange tissues.

Figure 5a. Ribonuclease protection analysis on RNA derived from two *Alicia* neonates (***A***), and a non-transgenic littermate (***NTG***). Samples collected were liver, brain with exchange tissue removed (***brain***), and the exchange tissues of the brain, including the choroid plexus (***CP***). RNA derived from these samples was hybridised with a *luciferase* probe and an *mGAP* probe, and protected fragments are visualised by autoradiography (see **Materials and Methods**). The two *Alicia* individuals assayed display a large variation in the level of *luciferase* expression in both ***liver*** and ***CP*** samples, but the reporter gene is expressed in these two tissues. *Luciferase* transcripts could not be detected in the depleted ***brain*** samples, or in samples derived from the non-transgenic littermate.

Figure 5b. Ribonuclease protection analysis on RNA derived from a single *Quasar* neonate (***Q***), and a non-transgenic littermate (***NTG***). The origin of the samples, and details of the assay are identical to those described in ***Figure 5a***. *Luciferase* mRNA can be detected in both ***liver*** and ***CP*** samples, but not in depleted ***brain*** samples from the transgenic individual. *Luciferase* message could not be detected in non-transgenic samples, or in a tRNA control.

Figure 5c. Ribonuclease protection analysis on RNA derived from two *Holly* neonates (***H***), and a non-transgenic littermate (***NTG***). The origin of the samples, and details of the assay are identical to those described in ***Figure 5a***. *Luciferase* mRNA can be detected in ***CP*** samples only, not in ***liver*** or depleted ***brain*** samples from the transgenic individual. *Luciferase* message could not be detected in non-transgenic samples, or in a tRNA control.

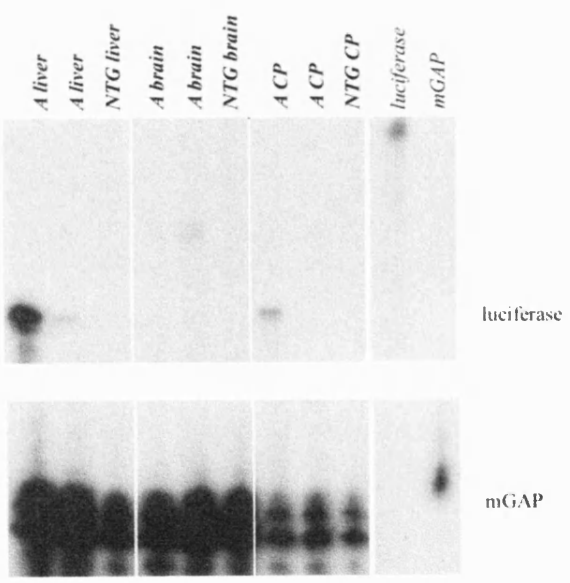


Figure 5a.

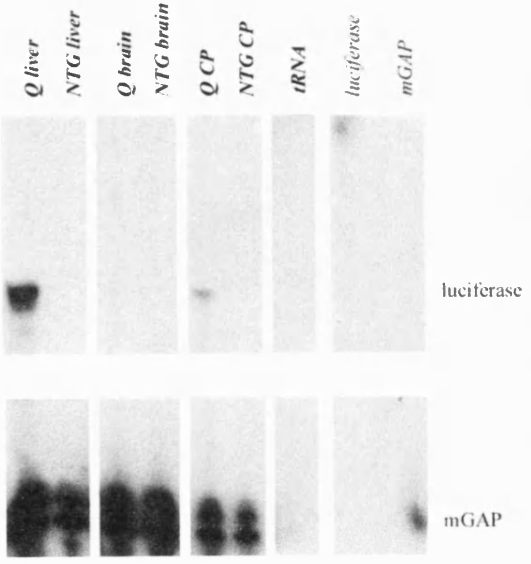


Figure 5b

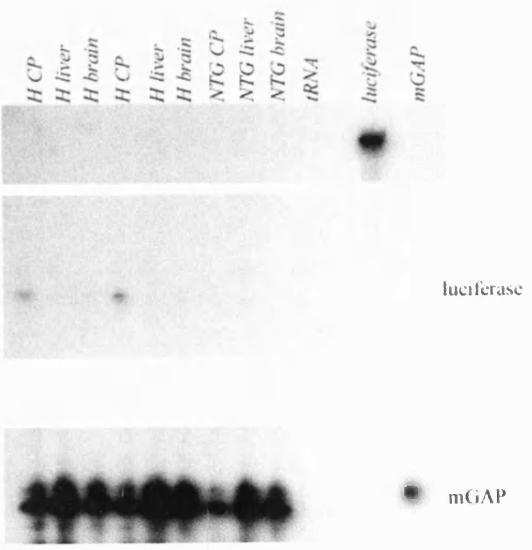


Figure 5c

Tissue-specific expression at e14.5

The data collected from D1 neonates has revealed tissue-specific expression patterns of the P3-*luciferase* transgenes, when *in-cis* to either the *H19* enhancers or the CCD. The *H19* enhancers drive expression in endodermally derived tissues, as well as in the brain, and the CCD drives reporter gene expression in the exchange tissues of the brain. Neither element could drive reporter gene expression to high levels in mesodermally derived tissues.

It next seemed appropriate to ask whether this expression was recapitulated at an earlier stage of development, in particular to corroborate the expression from CCD-containing transgenes in the brain. This brain expression has been shown to be confined to the choroid plexus and leptomeninges.

The choroid plexus first appears in the fourth ventricle of the developing mouse brain at e12, having arisen from the ependymal layer of the myelencephalon [202, 203]. By e13 the anterior choroid plexus has expanded laterally into the first two ventricles. By e15 there is a choroid plexus in the third ventricle. At the time point of embryonic day 14.5 the anterior and fourth ventricle choroid plexi should be fully formed, and the third ventricle choroid plexus in the process of forming. To allow the analysis of a large number of samples, a crude dissection was performed on e14.5 embryos, with the head taken to assess the level of reporter gene expression in the exchange tissues of the brain. It has been shown above that at D1 the exchange tissues constitute the majority of head expression driven by the CCD. There is very little expression from ***H*** transgenes in the remainder of the brain, or in tongue and skeletal muscle. It cannot be ruled out, however, that there may be other tissues in the head that can utilise the CCD to drive reporter gene expression above background levels.

The question of whether either the *H19* enhancers or the CCD can drive expression in extraembryonic tissues could not be tackled by the D1 analysis, as these tissues cannot be recovered post-partum. As discussed in the introduction, the mesodermal components of the chorioallantoic placenta and the yolk sac are both major sites of *Igf2* expression in mid-gestation mouse embryos (Lee, 1990). There is some evidence that the *H19* enhancers drive the expression of *Igf2* in the yolk sac [105], but not in the placenta [104].

Luciferase reporter gene activity has not previously been assayed in the extraembryonic tissues. The following experiments will test the contribution of the *H19* enhancers (***A***-lines), the CCD (***H***-lines) and both elements together (***Q***-lines) on the expression of *luciferase* from P3 in these tissues.

A further question in this section is do the two elements interact to modulate expression levels? At the D1 time-point, expression levels between lines with a particular construct were highly variable, preventing any comparison between constructs. Expression from the P3 promoter has been shown to decline shortly after birth both at its endogenous location [204], and in the transgenes utilised by this study [108]. Transgenic lines could vary slightly in the timing at which they begin to downregulate expression from the P3 promoter, resulting in large quantitative differences in expression levels between the lines. At mid-gestation, expression is predicted to be maximal, and therefore it follows that expression levels between lines might be more comparable. Body samples (whole embryo minus head) were collected primarily to test for quantitative differences in reporter gene expression levels arising from an association between the *H19* enhancers and the CCD.

As in the D1 analysis, tissues (head, body, yolk sac, and placenta) were collected from the offspring of crosses of females or males hemizygous for the transgene vs. F1 partners (the background strain, see **Materials and Methods**). Embryos were collected 14 days following the appearance of a copulation plug. Reporter gene expression levels were measured in tissues derived from transgenic mice bearing the ***A*** construct (P3-*luciferase*-*H19* enhancers), the lines *Alicia, Axe* and *Ayah*; the ***H*** construct (P3-*luciferase*-CCD), *Harold, Holly* and *Hamish* lines; and the ***Q*** construct (P3-*luciferase*-CCD-*H19* enhancers), *Quasar, Quark* and *Quiche* lines. A single line bearing P3-*luciferase* with no additional elements (the ***M*** construct, *Marcus*) was included, as levels of reporter gene expression from P3 alone have not previously been assayed at this stage. ***Figure 6*** shows tissue-specific reporter gene activity across the lines. Expression following male transmission and female transmission of the transgene are shown separately for each transgenic line, as in some cases the means are significantly different (see **Chapter 5**).

In all tissues examined, expression of the *luciferase* gene could not be detected above background levels in the *Marcus* line. The mean *luciferase* specific activity from *Marcus* samples is included below to represent the background specific activity in this assay.

Head expression (**Figure 6a**).

In head samples reporter gene activity is seen in samples representative of all three reporter constructs. The highest levels of gene activity were observed in those lines containing the CCD (*Quark, Quasar* and *Harold*).

Body expression (**Figure 6b**).

High levels of reporter gene expression in the bodies of transgenic mice were observed only when the *H19* enhancers were present. ***H*** lines all expressed the transgene at levels comparable to *Marcus.* ***Q*** lines expressed the transgene at ~5 fold greater levels than ***A*** lines (based on median figures for each construct).

Yolk-sac expression (**Figure 6c**).

In yolk sac, those lines containing the *H19* enhancers expressed the reporter gene above background levels, but at very low levels. ***A*** and ***Q*** lines expressed the transgene at approximately equal levels.

Placental expression (**Figure 6d**).

All of the transgenic lines assayed displayed reporter gene activity in the placenta that was very close to background levels.

Figure 6. Tissue-specific expression across the lines, at embryonic day 14.5.

Figure 6a. Mean reporter gene activity in head samples across the lines, following male (M) and female (F) transmission of the transgene. The highest levels of reporter gene activity are seen in ***Q*** and ***H*** lines, both of which contain the CCD. Error bars show the standard error of the mean.

Figure 6b. Mean reporter gene activity in body samples across the lines, following male (M) and female (F) transmission of the transgene. Reporter gene activity is seen in ***A*** and ***Q*** lines, both of which contain the *H19* enhancers. Error bars show the standard error of the mean.

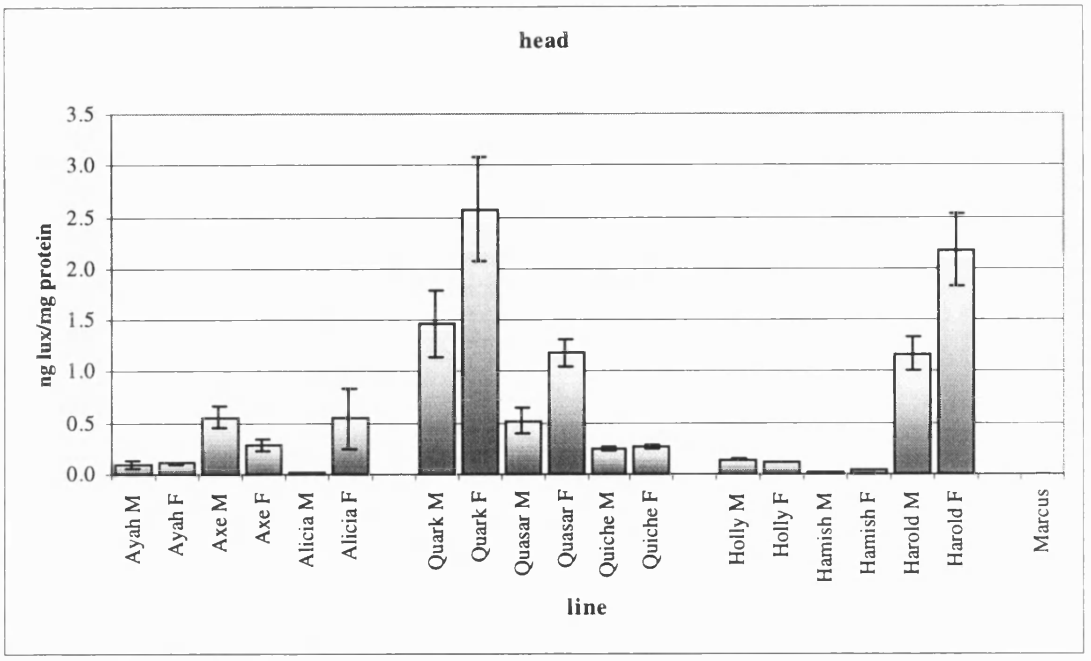


Figure 6a

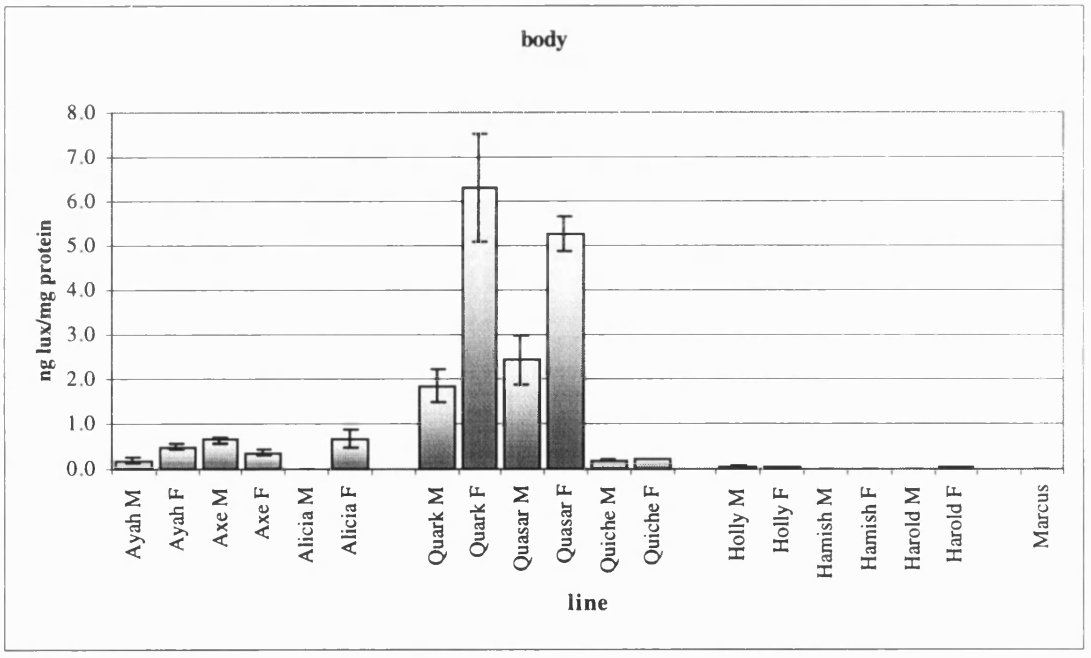


Figure 6b.

Figure 6 (cont.). Tissue-specific expression across the lines, at embryonic day 14.5.

Figure 6c. Mean reporter gene activity in yolk-sac samples across the lines, following male (M) and female (F) transmission of the transgene. Reporter gene activity is seen in those lines that contain the *H19* enhancers. Error bars show the standard error of the mean

Figure 6d. Mean reporter gene activity in yolk-sac samples across the lines, following male (M) and female (F) transmission of the transgene. Extremely low expression levels are seen in all lines, independent of reporter construct. Error bars show the standard error of the mean.

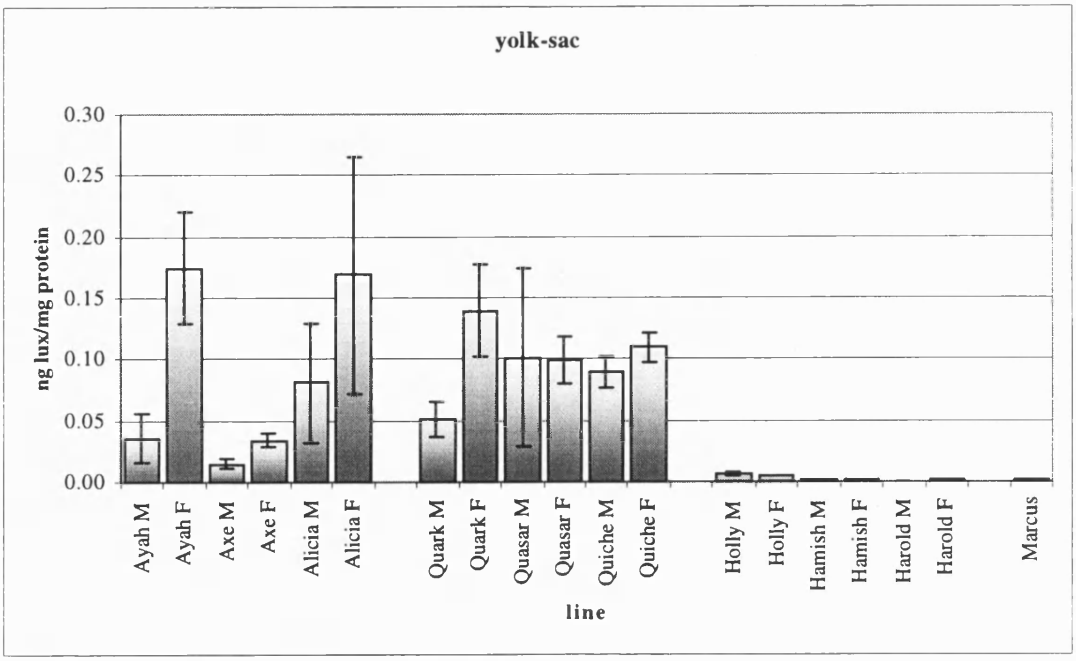


Figure 6c.

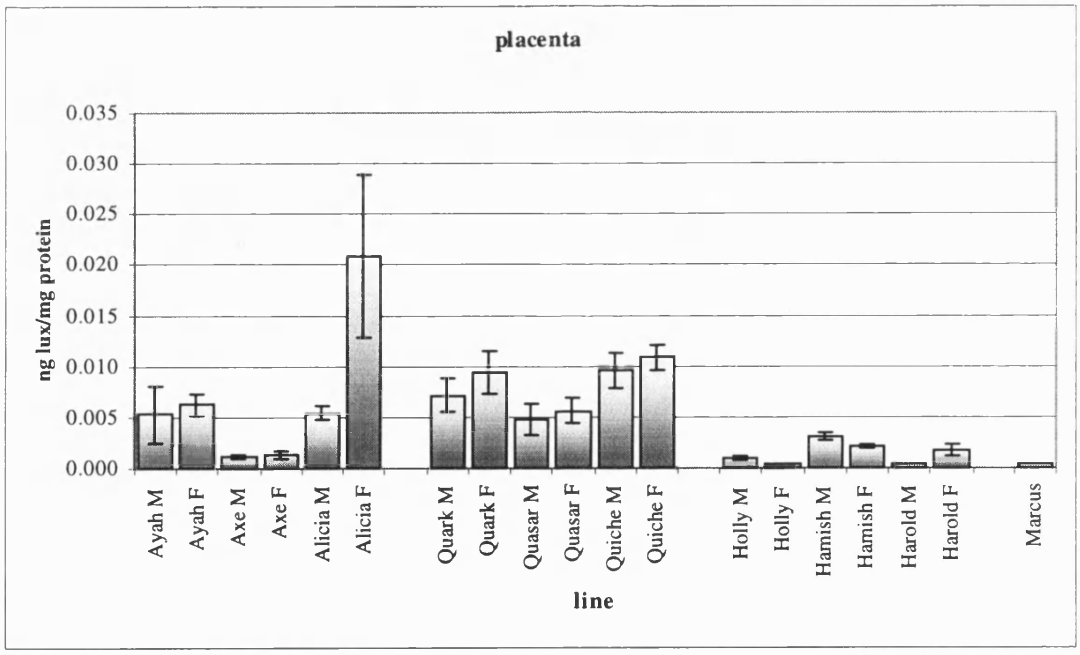


Figure 6d

Conclusions

The first question posed by this analysis was; what tissue specificity is conferred by the *H19* enhancers/CCD when linked to *luciferase*-P3? The CCD provided a good candidate for the yet undiscovered mesodermal enhancers at the *Igf2/H19* locus. A previous analysis has shown that the CCD cannot drive reporter gene expression above basal levels in cardiac mesoderm [108]. This result is unsurprising in the light of a study in which a YAC transgene containing ~130kb of genomic sequences including the *Igf2* and *H19* genes was shown not to express either *H19* or an *Igf2*-linked *β-galactosidase* reporter gene in heart tissue [36]. The YAC transgene did, however, recapitulate other mesoderm-specific expression patterns of these genes, demonstrating that some of the necessary regulatory elements do lie within the 130kb region. The experiments described in this chapter demonstrate that the CCD is unable to drive reporter gene expression in the mesodermally-derived tongue and skeletal muscle, and do not act on the *H19* enhancers to broaden their range of tissue-specificity to these tissues.

Extremely low levels of reporter gene expression were detected in the liver and kidney when P3-*luciferase* was *in-cis* to the CCD, confirming that it does not act as an enhancer in these tissues. Conversely, when the *H19* enhancers were present in the transgenic constructs, very high levels of *luciferase* specific activity was detected in liver, and moderate levels were detected in kidney (perhaps reflecting a smaller endodermal component to this organ). The *H19* enhancers are known to drive expression in endodermal tissues, so this result is expected [103, 104, 108, 109].

In brain samples, significant levels of reporter gene expression were observed in those transgenic lines containing the CCD. Analysis of *luciferase* mRNA expression by ribonuclease protection analysis confirmed that this expression was confined mainly to the exchange tissues, consistent with the known gene expression of *Igf2* [97]. *Luciferase* mRNA was detected in the exchange tissues of the brains of *A*-construct-bearing transgenic mice too, suggesting that the *H19*-enhancers are also able to drive gene expression in these tissues. Leptomeningial expression from transgenes containing a placental alkaline phosphatase reporter gene *in-cis* to the *H19* enhancers was reported recently [109]. These results are interesting in the context of a study of *Igf2* expression in the exchange tissues of the rat [102]. Choroid plexus expression was shown to be monoallelic from e13.5 to e15.5,

with a subsequent switch to biallelic expression from e18.5, after which expression of *Igf2* remained biallelic. Such a switch could imply that more than one element is responsible for *Igf2* expression in the exchange tissues of the brain.

It was of interest to discover whether the ***H*** transgenes were expressed concomitantly with the development of the choroid plexus. *Luciferase* specific activity assays were performed on transgenic mice at embryonic day 14.5. At this stage the anterior and fourth ventricle choroid plexi have been formed, and the third ventricle choroid plexus is beginning to delaminate from the ependymal layer of the pia mater [202, 203].

Two out of three of ***H*** transgenic lines (*Holly, Hamish*) expressed *luciferase* at above-basal levels in head samples at e14.5. This suggests that the CCD is active in the brain at developmental stages concomitant with the development of the choroid plexus. However, as whole head was assayed in order to allow the collection of a large number of samples, this is not the only conclusion that can be taken from this data. The CCD may well drive gene expression in other head tissues at this developmental stage. ***A*** lines also exhibit above-basal levels of reporter gene expression in the head. As with the ***H*** transgenes, gene expression from the ***H19*** enhancers cannot be localised to the exchange tissues of the brain, as these elements are likely to drive gene expression in other head tissues. Only *in-situ* expression data will be able to resolve expression patterns in the head driven by the CCD and the *H19* enhancers.

The analysis at e14.5 tested the possibility that the CCD or the *H19* enhancers might be responsible for the expression of *Igf2* in extraembryonic tissues. Both the allantoic component of the placenta, and the yolk sac [70, 87, 88] are major sites of *Igf2* gene expression at mid-gestation in the mouse. Transgenic mice carrying a construct composed of a ***H19*** minigene *in-cis* to the *H19* enhancers displayed transgene expression in yolk-sac in some lines, but not others [105]. Paternal deletion of the *H19* enhancers results in the reduction of *Igf2* expression in the yolk sac by 30% compared to wild-type levels [104]. In this study the *H19* enhancers in concert with P3-*luciferase* demonstrated above-background levels of reporter gene expression in yolk sac at e14.5. The *luciferase* mean specific activities were at ~10 times lower levels than found in the embryonic tissues, suggesting that these enhancers may not account for the totality of expression in this tissue where the *Igf2* gene is expressed abundantly [88]. *In-situ* studies could elucidate whether the expression from the *H19* enhancers is distributed to all, or a subset of cells within this tissue. The CCD made no contribution to expression of the reporter gene in the yolk sac.

In the placenta there was no significant upregulation of *luciferase* gene expression from P3 in any transgenic line. It must be concluded that neither the *H19* enhancers nor the CCD can drive reporter gene expression in this organ at e14.5. This result is unsurprising, since a 130kb YAC extending from promoter 1 of *Igf2* to 35kb downstream of *H19* does not express the *H19* gene, or an *Igf2*-linked *Lac-Z* reporter gene in placenta (J. Ainscough, pers. comm.). There is accumulating evidence that the enhancers for *Igf2* (and possibly *H19*) in this tissue lie upstream of *Igf2*. P3-*luciferase* transgenes containing a 2kb region including DMR1 (***T*** lines, ***Figures 1*** and ***3***) express the reporter gene above background levels in the placenta at e14.5 (G. Dell and M. Charalambous, unpublished). The upstream region is also implicated in the production of a placental-specific transcript, and placental-specific methylation [6, 14].

A second question within this analysis was; do the *H19* enhancers and the CCD interact to create new patterns of gene expression, or modulate expression levels? In transgenic mice at D1, no new patterns of gene expression were revealed in ***Q*** construct bearing transgenic mice, as compared with ***A*** lines or ***H*** lines, at least in the limited number of tissues analysed. It was impossible to compare quantitative expression levels resulting from construct type, due to the variability of expression between lines bearing the same construct (e.g., *Quark* vs. *Quasar*). This variability may be due to one or more of the transgenes being subject to position effects, i.e., the low-expressing *Quark* and *Hamish* lines could have the transgene inserted into a heterochromatic region of the genome, thus silencing the *luciferase* gene. For *Quark*, this explanation seems unlikely, as liver expression is still very high, and the transgene has been shown to be unmethylated in liver, brain and tongue (see next chapter). Another explanation for discrepancies between quantitative gene expression levels for transgenic lines with the same construct comes from the observation that both *in-vitro* [204] and *in-vivo* [108], expression from the P3 promoter has been shown to decline shortly after birth. Transgenic lines could vary slightly in the timing at which they begin to downregulate expression from the P3 promoter, resulting in large quantitative differences in expression levels between the lines. At an earlier developmental stage, when such a shift in promoter activity is not expected, transgene expression levels might be more uniform. Indeed, *luciferase* specific activities at e14.5 were found to be more uniform for lines derived from a particular transgene construct, within each set of tissues assayed.

Comparisons of quantitative expression levels of ***A*** lines vs. ***Q*** lines in body samples revealed that ***Q*** lines show an ~5-fold greater level of *luciferase* specific activity than ***A*** lines. ***H*** lines are not expressed above background levels in the body.

In some cases ([205], [206] and reviewed in [207]) transgene expression levels can be directly correlated to copy number, but only in the presence of a locus control region (LCR). When the LCR is deleted from these transgenes, this correlation is lost. Other groups report no correspondence between transgene copy number and expression levels [208]. As shown in ***Appendix 1***, there is no correlation between transgene copy number and levels of reporter gene expression in this case.

It is possible that transgene size may play a role in the level of gene expression, perhaps larger transgenes insulate their reporter genes more effectively from flanking genomic DNA, or stabilise the reporter gene in an 'open' chromatin conformation. The ***E*** transgene (***Figure 3*, Chapter 1**) is of approximately equal size to the ***Q*** transgene, containing a ~2kb DMR1 region instead of the CCD. Mice bearing ***E*** transgenes express *luciferase* at comparable levels to ***A*** in the body at e14.5 (not shown), suggesting that transgene size alone is not a factor here. The DMR1, however, is not a 'neutral' region of DNA, and its effect here may be misleading.

Without *in-situ* data it is impossible to distinguish whether this synergistic effect between the two elements seen in ***Q*** lines is due to expression of *luciferase* in a broader range of tissues, or to increased enhancer activity within the same tissues.

Quantitative assays of *luciferase* reporter gene expression in transgenic mice bearing the ***A*** construct (*H19* enhancers), the ***H*** construct (CCD) and the ***Q*** construct (both elements), at two developmental stages, D1 and e14.5 have revealed:

Neither the CCD nor the *H19* enhancers can drive reporter gene expression to high levels in the mesodermally derived tissues skeletal muscle and tongue, or in the placenta.

The *H19* enhancers in P3-*luciferase* transgenics act in the expected manner, by driving gene expression in the endodermally derived liver and kidney. These enhancers were also shown to drive gene expression in the yolk sac of the e14.5 embryo.

The CCD can drive reporter gene activity in brain at D1, specifically in the exchange tissues, and may be contributing to expression in the head at e14.5.

At e14.5 the two elements manifest a synergistic effect, with expression levels from ***Q*** lines exceeding expression levels from ***A*** lines by approximately 5-fold, in the embryonic

tissues. The significance of this small effect awaits verification by the study of *in-situ* patterns of transgene expression.

CHAPTER 5: PARENTAL ORIGIN-SPECIFIC EFFECTS OF *P3-LUCIFERASE* TRANSGENES.

Introduction

There is increasing evidence that the *H19* enhancers play a minor role in the imprinting mechanism of *Igf2* and *H19*. Transgenes containing a *H19* mini-gene with 5kb upstream sequence and 10kb downstream sequence, including these enhancers were found to be reproducibly silenced when inherited paternally, and expressed the mini-gene when inherited maternally, at ectopic locations in the mouse genome [37]. These transgenes also exhibited methylation patterns in the 5' flanking region and within the gene itself, which were very similar to those observed at the endogenous locus. Transgenes derived from these, that did not contain the *H19* enhancers were found not to be expressed, and furthermore, the parent-of-origin-specific methylation patterns of the transgene were lost. Two conclusions could be drawn from this work; i) expression of a gene is required for its correct methylation/imprinting, or ii) sequences are present in the enhancer region that stabilise the imprinting process. The second explanation is supported by the work of Ward et. al. [108], in which P3-*luciferase* reporter transgenes containing the *H19* enhancers regularly displayed higher levels of *luciferase* activity when the transgene was inherited maternally, than when it was paternally inherited. Though this was a minor effect (2-3-fold in most cases), this, and the observation that only very large [36], or very high copy number ([38, 39, 37, 156]) transgenes are imprinted, suggests that imprinting at this region could be a result of multiple factors, some with minor activities.

Choroid plexus and meningial expression of *Igf2* is biallelic in several mammalian species including mouse [98] and rat [97]. Elements that drive expression in these tissues must somehow overcome or bypass the imprinting mechanism. As a candidate for an enhancer element in these tissues, the CCD may have properties that allow it to moderate imprinting effects. No such parent-of-origin-specific differences have yet been observed in transgenic mice carrying the CCD, so the possibility exists that this element can induce loss of imprinting in transgenes.

Can either the *H19* enhancers or the CCD confer parent-of-origin specific differences in gene expression? As discussed above, this question has been answered in

part by a previous study [108], in which parent-of-origin effects were observed in the livers of those lines bearing the ***A*** construct. In this case, a 2-3-fold increase in expression was observed when the transgene was inherited maternally above levels of expression from paternal transgene transmission. Patterns of ***H*** construct expression appeared to be consistent with *Igf-2* expression, i.e., when expressed in the brain the reporter gene showed no parental bias in gene expression.

This study extends the previous analysis into additional tissues, and examines if the two elements combined increase or diminish parent-of-origin specific effects.

Results

Parental origin specific effects at D1

As described in **Chapter 4**, tissue samples for the D1 analysis of transgene expression were collected from the offspring of both male and female hemizygotes. This allowed a test for expression differences when the transgene was transmitted through the male or female germline. At least 20 individuals of each transgenic line, following male or female transmission of the transgene, were collected to allow comparisons of mean *luciferase* specific activity by a Student's *t* test. Expression levels were measured in tissues derived from transgenic mice bearing the ***A*** construct (P3-*luciferase-H19* enhancers), *Alicia, Archy* and *Ayah* lines; the ***H*** construct (P3-*luciferase*-CCD), the *Harold, Holly* and *Hamish* lines; and the ***Q*** construct (P3-*luciferase*-CCD-*H19* enhancers), the *Quasar, Quark* and *Quiche* lines (all shown in ***Figure 3***, **Chapter 1**). Expression levels were examined in three transgenic lines representing each construct in an attempt to distinguish real effects on P3 due to the element of interest from position effects caused by transgene integration. Any effects observed across a majority of lines bearing a particular element will be regarded as real.

Of particular interest will be any test that shows a *p*-value of less that 0.01, when there is a >2-fold difference between the means. These conditions are imposed arbitrarily, with the aim of distinguishing true parental origin specific effects from Type I errors. This data is summarised in ***Table 6*** and represented graphically in ***Figures 7, 8*** and ***9***.

Line: χ values ng lux/mg protein	*Quark*	*Quasar*	*Quiche*	*Archy*	*Alicia*	*Ayah*	*Holly*	*Hamish*	*Harold*
Brain M	$1.69x10^{-2}$	0.72	$3.16x10^{-1}$	$1.52x10^{-2}$	$1.21x10^{-2}$	$2.36x10^{-3}$	0.29	$2.39x10^{-2}$	0.14
F	$1.81x10^{-2}$	0.85	$3.15x10^{-1}$	$1.35x10^{-2}$	$5.93x10^{-2}$	$1.58x10^{-3}$	0.32	$1.74x10^{-2}$	0.15
p-value:	*0.049*	*0.124*	*0.987*	*0.500*	*0.299*	*0.019*	*0.408*	*0.0007*	*0.637*
TongueM	$2.10x10^{-4}$	$8.84x10^{-3}$	$6.58x10^{-2}$	$4.72x10^{-2}$	$1.82x10^{-2}$	$6.54x10^{-2}$	$6.80x10^{-3}$	$3.20x10^{-4}$	$9.86x10^{-2}$
F	$4.58x10^{-4}$	0.12	$1.60x10^{-1}$	$8.43x10^{-3}$	$4.61x10^{-2}$	$1.07x10^{-1}$	$1.16x10^{-2}$	$3.61x10^{-4}$	$9.63x10^{-2}$
p-value:	***4.11x10^{-3}***	***3.82x10^{-6}***	*0.027*	*0.126*	*0.049*	*0.073*	*0.298*	*0.635*	*0.912*
Muscle M	$4.19x10^{-4}$	$2.82x10^{-2}$	$1.03x10^{-1}$	$7.21x10^{-2}$	$4.40x10^{-3}$	$8.33x10^{-2}$	$4.12x10^{-2}$	$1.10x10^{-4}$	$3.39x10^{-2}$
F	$5.44x10^{-4}$	$2.49x10^{-2}$	$1.82x10^{-2}$	$8.15x10^{-3}$	$5.99x10^{-3}$	$5.03x10^{-2}$	$2.71x10^{-2}$	$1.61x10^{-4}$	$1.75x10^{-2}$
p-value:	*0.629*	*0.754*	*0.340*	***8.15x10^{-3}***	*0.671*	*0.383*	*0.613*	*0.276*	*0.016*
Liver M	2.68	10.41	14.12	25.66	4.74	7.52	$4.00x10^{-4}$	$5.10x10^{-6}$	$3.00x10^{-3}$
F	5.69	15.21	18.49	6.47	54.40	1.84	$6.21x10^{-4}$	$7.14x10^{-6}$	$3.44x10^{-4}$
p-value:	***3.31x10^{-3}***	*0.108*	*0.182*	***3.70x10^{-3}***	***6.38x10^{-3}***	*0.013*	*0.087*	*0.348*	*0.009*
KidneyM	$2.69x10^{-3}$	$2.80x10^{-2}$	$2.50x10^{-2}$	$8.76x10^{-2}$	$2.26x10^{-2}$	$8.70x10^{-2}$	$2.40x10^{-3}$	$1.79x10^{-5}$	$1.00x10^{-4}$
F	$4.95x10^{-3}$	$4.91x10^{-2}$	$2.26x10^{-2}$	$7.67x10^{-3}$	$6.22x10^{-2}$	$7.76x10^{-2}$	$1.56x10^{-3}$	$2.53x10^{-5}$	$2.03x10^{-4}$
p-value:	*0.149*	*0.135*	*0.848*	***7.83x10^{-3}***	*0.088*	*0.694*	*0.063*	*0.111*	*0.121*

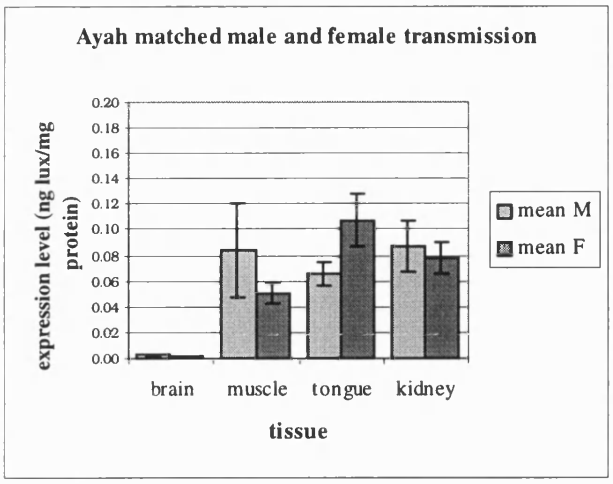
Table 6. Summary of expression data of the *luciferase* reporter constructs across the transgenic lines, at D1. Levels of expression following male (M) and female (F) transmission of the transgene are shown as mean *luciferase* specific activity (ng *luciferase*/mg soluble protein. Differences in the means of each tissue in each line following paternal or maternal transmission are compared using a Student's *t*-test, with the null hypothesis; 'there is no difference between the means following maternal or paternal transmission of the transgene'. The *p*-value for each comparison is shown, and highlighted in light grey if a significant difference between the means is observed, i.e. if *p* is in the range 0.01 -0.05, or the difference between the means is <2-fold at a high level of significance $p<0.01$. Darker grey boxes represent those tests where a highly significant difference was found (i.e. $p<0.01$) between the means, with a greater than 2-fold difference. The tests highlighted in dark grey represent those tests that conform to the conditions in which the null hypothesis will be rejected.

Figure 7. Parent of origin-specific effects at D1, the ***A*** construct.

Figure 7a. *Ayah* matched male and female transmission. Mean reporter gene expression levels (ng *luciferase*/mg protein) are shown for each tissue following male (M) and female (F) transmission of the transgene. Liver expression is shown on a separate axis due to a large difference in scale. Liver samples show a low, but significant difference in expression levels following maternal vs. paternal transmission of the transgene ($p=0.13$), paternal highest. Error bars show the standard error of the mean in each case.

Figure 7b. *Archy* matched male and female transmission. Mean reporter gene expression levels (ng *luciferase*/mg protein) are shown for each tissue following male (M) and female (F) transmission of the transgene. Liver expression is shown on a separate axis due to a large difference in scale. Muscle ($p=8.2x10^{-3}$), kidney ($p=7.8\ x10^{-3}$) and liver ($p=8.2x10^{-3}$) samples all show a significant difference in expression levels following maternal vs. paternal transmission of the transgene, paternal highest. Error bars show the standard error of the mean in each case.

Figure 7c. *Alicia* matched male and female transmission. Mean reporter gene expression levels (ng *luciferase*/mg protein) are shown for each tissue following male (M) and female (F) transmission of the transgene. Liver expression is shown on a separate axis due to a large difference in scale. Tongue ($p=0.049$) and liver ($p=6.38\ x10^{-4}$) samples show a significant difference in expression levels following maternal vs. paternal transmission of the transgene, maternal highest. Error bars show the standard error of the mean in each case.



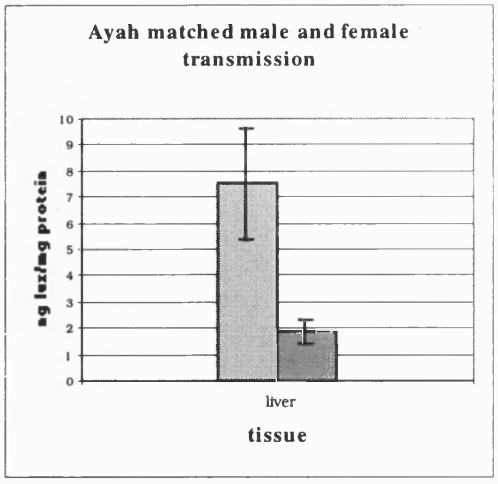
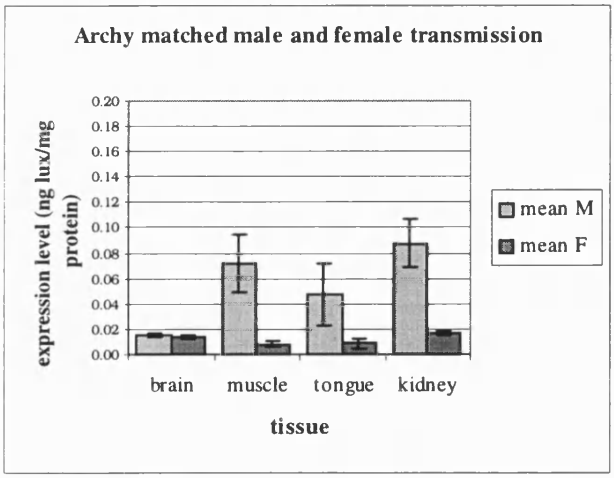


Figure 7a.



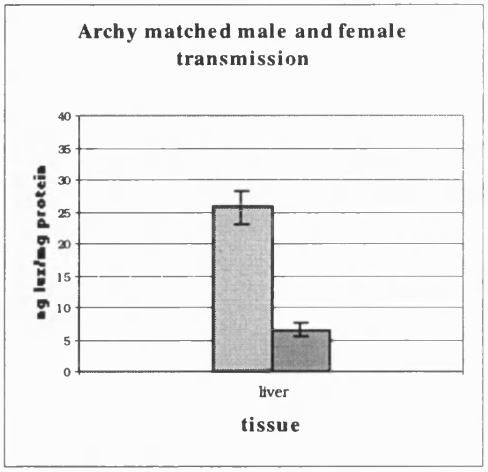
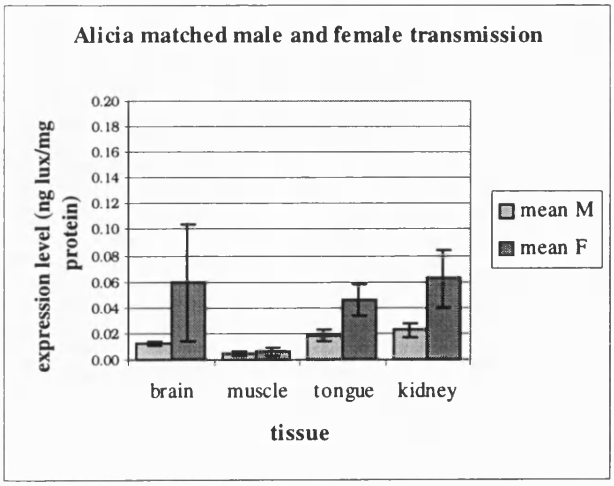


Figure 7b.



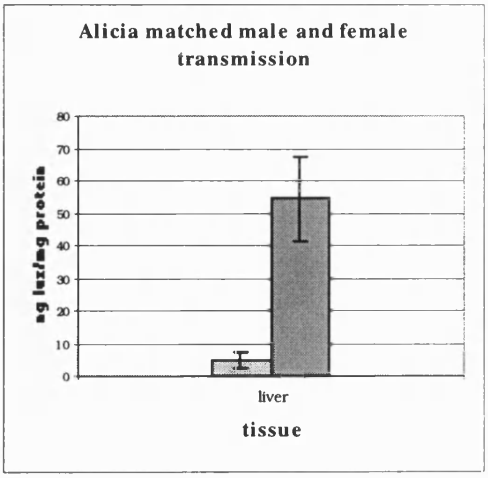


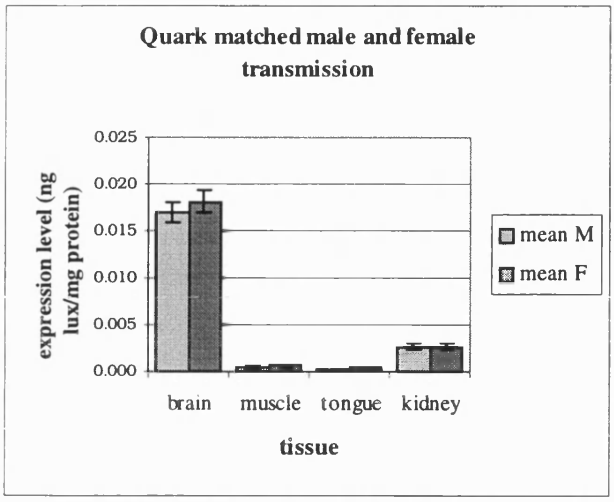
Figure 7c.

Figure 8. Parent of origin-specific effects at D1, the ***Q*** construct.

Figure 8a. *Quark* matched male and female transmission. Mean reporter gene expression levels (ng *luciferase*/mg protein) are shown for each tissue following male (M) and female (F) transmission of the transgene. Liver expression is shown on a separate axis due to a large difference in scale. Tongue ($p=4.1x10^{-3}$) and liver ($p=3.3x10^{-4}$) samples show a significant difference in expression levels following maternal vs. paternal transmission of the transgene, maternal highest. Error bars show the standard error of the mean in each case.

Figure 8b. *Quasar* matched male and female transmission. Mean reporter gene expression levels (ng *luciferase*/mg protein) are shown for each tissue following male (M) and female (F) transmission of the transgene. Liver expression is shown on a separate axis due to a large difference in scale. Liver ($p=3.8x10^{-6}$) samples show a significant difference in expression levels following maternal vs. paternal transmission of the transgene, maternal highest. Error bars show the standard error of the mean in each case.

Figure 8c. *Quiche* matched male and female transmission. Mean reporter gene expression levels (ng *luciferase*/mg protein) are shown for each tissue following male (M) and female (F) transmission of the transgene. Liver expression is shown on a separate axis due to a large difference in scale. Tongue ($p=0.027$) samples show a significant difference in expression levels following maternal vs. paternal transmission of the transgene, maternal highest. Error bars show the standard error of the mean in each case.



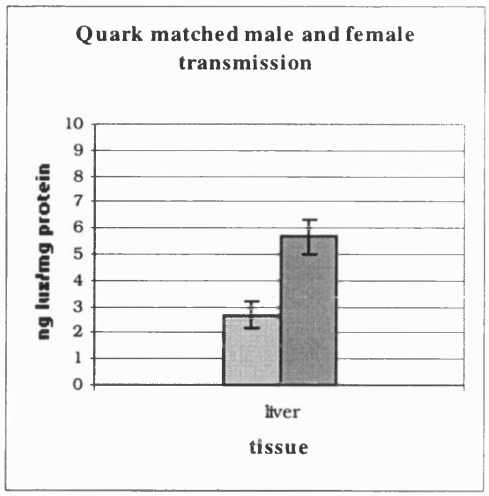
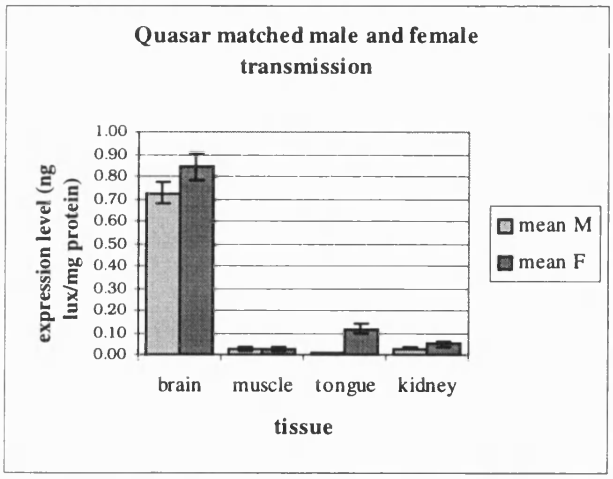


Figure 8a.



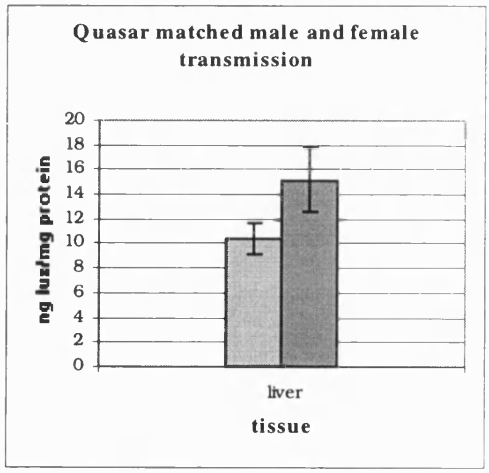
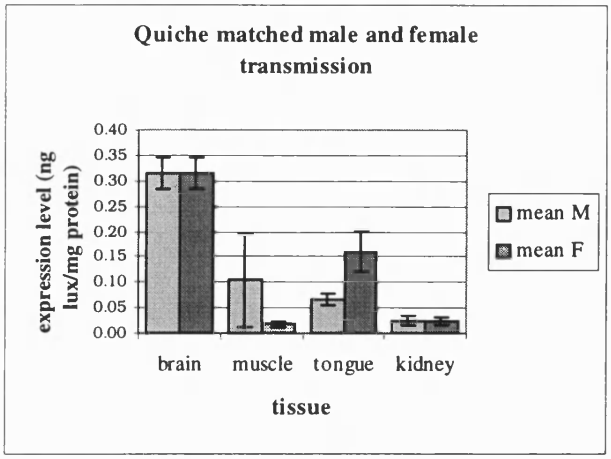


Figure 8b.



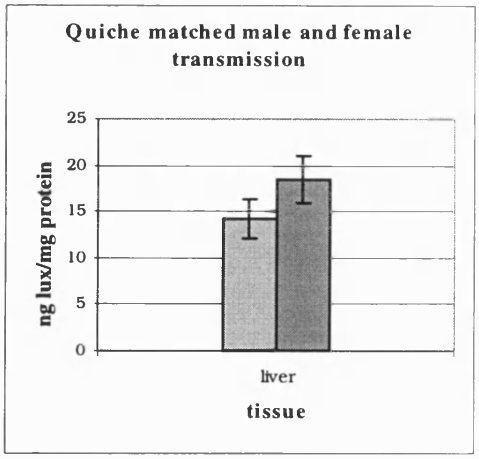


Figure 8c.

Figure 9. Parent of origin-specific effects at D1, the ***H*** construct.

Figure 9a. *Hamish* matched male and female transmission. Mean reporter gene expression levels (ng *luciferase*/mg protein) are shown for each tissue following male (M) and female (F) transmission of the transgene. Brain samples show a significant difference in expression levels following maternal vs. paternal transmission of the transgene, paternal highest, but the difference is less than two-fold ($p=7.0x10^{-4}$). Error bars show the standard error of the mean in each case.

Figure 9b. *Harold* matched male and female transmission. Mean reporter gene expression levels (ng *luciferase*/mg protein) are shown for each tissue following male (M) and female (F) transmission of the transgene. Muscle ($p=0.016$) and liver ($p=9.2x10^{-3}$) samples show a significant difference in expression levels following maternal vs. paternal transmission of the transgene, paternal highest. Error bars show the standard error of the mean in each case.

Figure 9c. *Holly* matched male and female transmission. Mean reporter gene expression levels (ng *luciferase*/mg protein) are shown for each tissue following male (M) and female (F) transmission of the transgene. Error bars show the standard error of the mean in each case. There are no significant differences in expression between reporter genes inherited maternally compared to reporter genes inherited paternally.

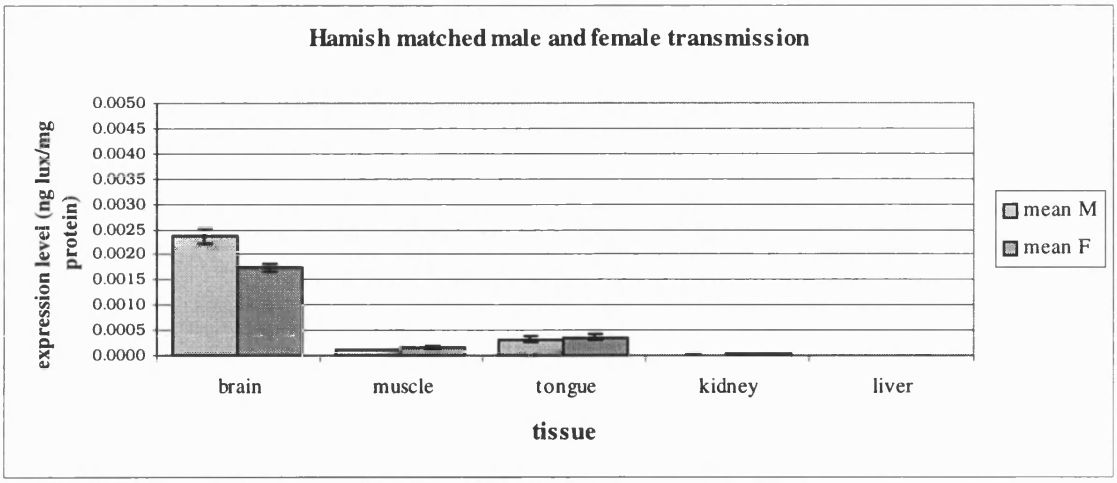


Figure 9a.

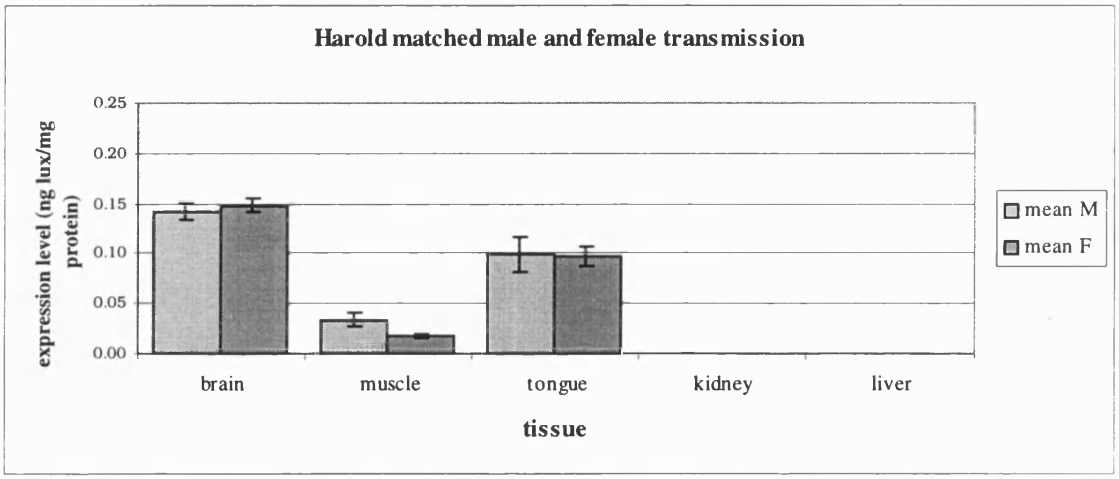


Figure 9b.

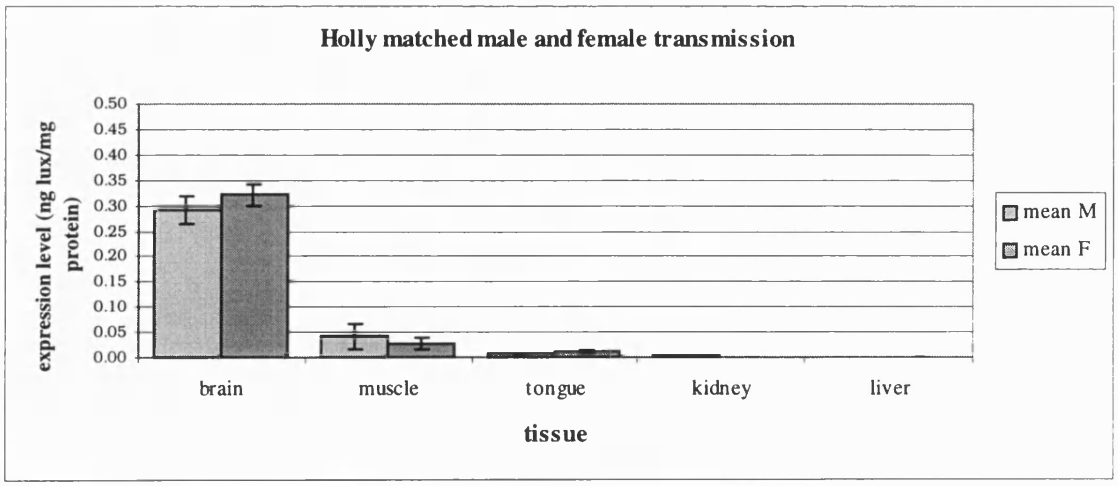


Figure 9c.

A lines (***Figure 7***)

Samples derived from the *Ayah* line show parent-of-origin effects in the liver at a low level of significance (p=0.013), where reporter gene expression is ~4-fold higher following paternal transmission of the transgene.

Samples derived from the *Archy* line show parent-of-origin specific effects in muscle, liver and kidney. All reporter gene expression is higher following paternal transmission of the transgene. In muscle reporter gene expression is ~9-fold higher ($p=8.2x10^{-3}$), in liver reporter gene expression is ~4-fold higher ($p=3.7x10^{-8}$) and in kidney reporter gene expression is ~11-fold higher ($p=7.8\ x10^{-3}$).

Samples derived from the *Alicia* line show parent-of-origin specific effects in the tongue and the liver. In tongue reporter gene expression is ~2.5-fold higher following maternal transmission of the transgene, at a low level of significance ($p=0.049$). In liver reporter gene expression is ~11-fold higher following maternal transmission of the transgene ($p=6.38\ x10^{-4}$).

Q lines (***Figure 8***)

Samples derived from the *Quark* line show an increase in expression levels following maternal transmission of the transgene in tongue and liver. In tongue maternal reporter gene expression is ~2-fold higher ($p=4.1x10^{-3}$) and in liver maternal reporter gene expression is also ~2-fold higher ($p=3.3x10^{-4}$) than paternal reporter gene expression. In the *Quasar* line the only parent-of-origin specific differences are observed in tongue where expression levels following maternal transmission are ~14-fold higher than paternal reporter gene expression ($p=3.8x10^{-6}$). The *Quiche* line also displays a parental origin-specific effect in the tongue, where expression levels following maternal transmission of the transgene are ~2-fold higher ($p=0.027$) than paternal gene expression levels.

H lines (***Figure 9***)

In all three ***H*** lines examined, only one parent-of-origin specific difference was observed in which gene expression following transmission from one parent was more than 2-fold greater than expression following transmission from the other parent. The single parental origin specific effect was seen in *Harold* liver, where despite very low expression levels paternal expression of the transgene is ~10-fold higher than maternal expression levels ($p=0.009$).

To summarise; the *H19* enhancers in concert with P3 appear to be able to confer parent-of-origin specific differences in *luciferase* gene expression in a subset of tissues. *Quark*, *Quasar*, *Quiche* and *Alicia* lines all show higher levels of reporter gene expression when the transgene is inherited maternally. In contrast, the *Archy* and *Ayah* lines show higher levels of expression when the transgene is inherited paternally. This result is contradictory to the previous analysis [108], where all expression differences were higher following maternal transmission of the transgene. The CCD is never able to create a parental bias in expression greater that 2-fold, and is unlikely to be an imprinting control element, unless it exerts an influence via other elements not tested in this study. As all three ***Q*** lines presented here show parent-of-origin-specific effects, the CCD does not appear to be overriding the imprinting signal, at least in the tissues assayed.

Parental origin specific effects at e14.5

Are the parental origin specific effects observed at D1 also present at earlier developmental stages? It was observed in the previous chapter that expression from the P3 promoter in *luciferase* transgenes is downregulated around the time of birth, to a variable degree between transgenic lines, making comparisons between lines problematic. This variable downregulation of P3 may complicate the analysis of parental origin specific effects.

Samples for the spatial analysis of transgene expression at e14.5 were collected from the offspring of both male and female hemizygotes. As with the D1 study, at least 20 individuals were collected to represent each parental cross, for each transgenic line, to allow comparison of mean *luciferase* expression in each tissue by a Student's *t*-test. The lines analysed were ***A*** lines (*Axe, Alicia* and *Ayah*), ***Q*** lines (*Quark, Quasar* and *Quiche*), and ***H*** lines (*Harold, Hamish* and *Holly*). At least 2 lines bearing each construct have also been analysed at D1 (see above).

As before, of interest are those lines that show at least a 2-fold difference between the means when the transgene is maternally vs. paternally inherited, at a significance of $p<0.01$. Mean *luciferase* specific activity following each parental transmission, and the *p*-value obtained from a Student's *t*-test are tabulated for each line in ***Table 7***, and represented graphically in ***Figures 10, 11*** and ***12***.

Line: χ values ng lux/mg protein	*Quark*	*Quasar*	*Quiche*	*Axe*	*Alicia*	*Ayah*	*Holly*	*Hamish*	*Harold*
Head M	1.470	$5.17x10^{-1}$	$2.47x10^{-1}$	$5.58x10^{-1}$	$1.23x10^{-2}$	$2.04x10^{-1}$	$1.36x10^{-1}$	$1.73x10^{-2}$	1.163
F	2.576	1.180	$2.61x10^{-1}$	$2.86x10^{-1}$	$5.46x10^{-1}$	$2.15x10^{-1}$	$1.84x10^{-1}$	$4.01x10^{-2}$	2.176
p-value:	*0.065*	***3.32x10^{-4}***	*0.651*	*0.027*	*0.054*	*0.889*	*0.074*	***9.37x10^{-4}***	*0.013*
Body M	1.842	2.425	$1.92x10^{-1}$	$6.39x10^{-1}$	$1.07x10^{-2}$	$3.77x10^{-1}$	$5.90x10^{-2}$	$2.36x10^{-3}$	$1.44x10^{-2}$
F	6.304	5.258	$2.09x10^{-1}$	$3.61x10^{-1}$	$6.59x10^{-1}$	$9.67x10^{-1}$	$4.32x10^{-2}$	$2.22x10^{-3}$	$4.43x10^{-2}$
p-value:	***9.14x10^{-4}***	***5.79x10^{-3}***	*0.557*	*0.006*	***7.1[illegible]***	***6.82x10^{-4}***	*0.074*	*0.432*	***5.01x10^{-4}***
Placenta M	$7.14x10^{-3}$	$4.72x10^{-3}$	$9.60x10^{-3}$	$1.13x10^{-3}$	$5.44x10^{-3}$	$1.06x10^{-2}$	*$9.72x10^{-4}$*	$1.94x10^{-3}$	$3.95x10^{-4}$
F	$9.40x10^{-3}$	$5.56x10^{-3}$	$1.09x10^{-2}$	$1.34x10^{-3}$	$2.09x10^{-2}$	$1.25x10^{-2}$	$3.96x10^{-4}$	$1.66x10^{-3}$	$1.74x10^{-3}$
p-value:	*0.374*	*0.654*	*0.559*	*0.569*	*0.054*	*0.742*	*0.001*	*0.042*	*0.033*
Yolk sac M	$5.09x10^{-2}$	$1.01x10^{-1}$	$8.97x10^{-2}$	$1.51x10^{-2}$	$8.08x10^{-2}$	$3.59x10^{-2}$	$6.83x10^{-3}$	$2.99x10^{-3}$	$5.89x10^{-4}$
F	$1.40x10^{-1}$	$9.93x10^{-2}$	$1.09x10^{-1}$	$3.38x10^{-2}$	$1.69x10^{-1}$	$1.75x10^{-1}$	$4.66x10^{-3}$	$2.07x10^{-3}$	$1.57x10^{-4}$
p-value:	*0.034*	*0.980*	*0.275*	*0.013*	*0.395*	***7.32x10^{-3}***	*0.237*	*0.015*	***3.60x10^{-3}***

Table 7. Summary of expression data of the *luciferase* reporter constructs across the transgenic lines, at e14.5. Levels of expression following male (M) and female (F) transmission of the transgene are shown as mean *luciferase* specific activity (ng *luciferase*/mg soluble protein). Differences in the means of each tissue in each line following paternal or maternal transmission are compared using a Student's *t*-test, with the null hypothesis; 'there is no difference between the means following maternal or paternal transmission of the transgene'. The *p*-value for each comparison is shown, and highlighted in light grey if a significant difference between the means is observed, *p* is in the range 0.01-0.05, or the difference between the means is less than 2-fold at a high level of significance (i.e., $p<0.01$). Darker grey boxes represent those tests where a highly significant difference (i.e. *p* is less than 0.01) was found between the means that was greater than 2-fold. The tests highlighted in dark grey represent those tests that conform to the conditions in which the null hypothesis will be rejected.

Figure 10. Parental origin-specific effects at e14.5, ***A*** construct

Figure 10a. *Alicia* matched male and female transmission. Mean reporter gene expression levels (ng *luciferase*/mg protein) are shown for each tissue following male (M) and female (F) transmission of the transgene. Body samples show a significant difference in expression levels following maternal vs. paternal transmission of the transgene, maternal highest ($p=7.13x10^{-4}$). Error bars show the standard error of the mean in each case.

Figure 10b. *Axe* matched male and female transmission. Mean reporter gene expression levels (ng *luciferase*/mg protein) are shown for each tissue following male (M) and female (F) transmission of the transgene. Head ($p=0.027$), body ($p=0.006$) and yolk sac ($p=0.013$) samples show a significant difference in expression levels following maternal vs. paternal transmission of the transgene, paternal highest, but with a less than two-fold difference between the means in head and body samples. Error bars show the standard error of the means.

Figure 10c. *Ayah* matched male and female transmission. Mean reporter gene expression levels (ng *luciferase*/mg protein) are shown for each tissue following male (M) and female (F) transmission of the transgene. Body ($p=6.82x10^{-4}$) and yolk sac ($p=7.32x10^{-3}$) samples show a significant difference in expression levels following maternal vs. paternal transmission of the transgene, maternal highest. Error bars show the standard error of the mean in each case.

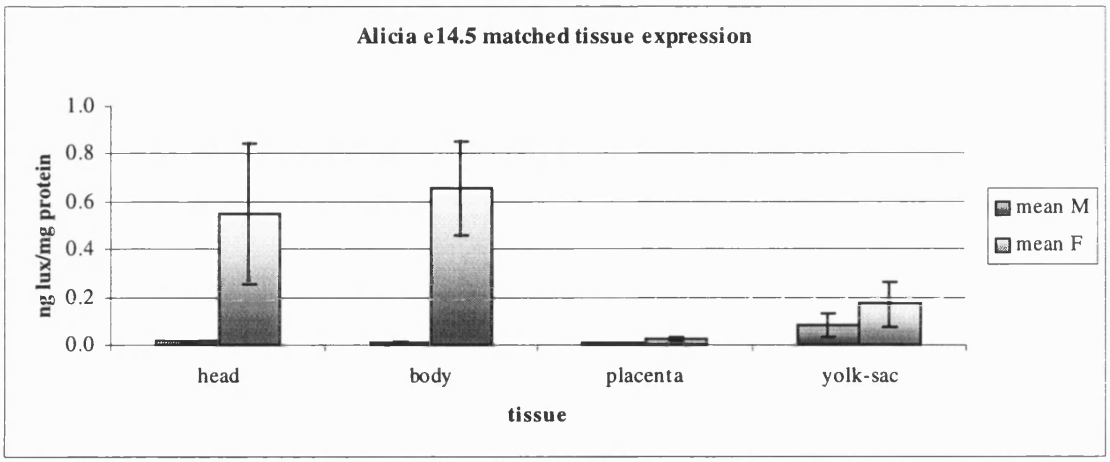


Figure 10a.

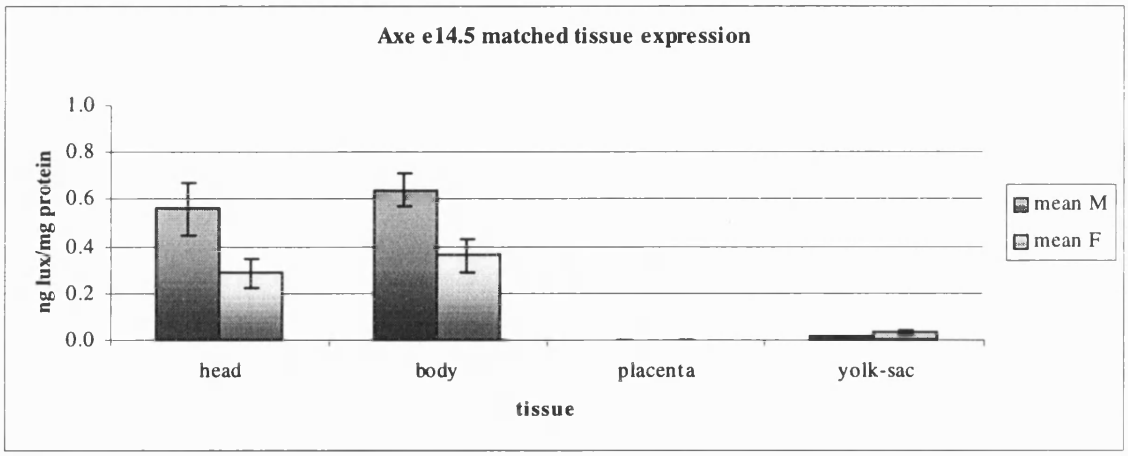


Figure 10b.

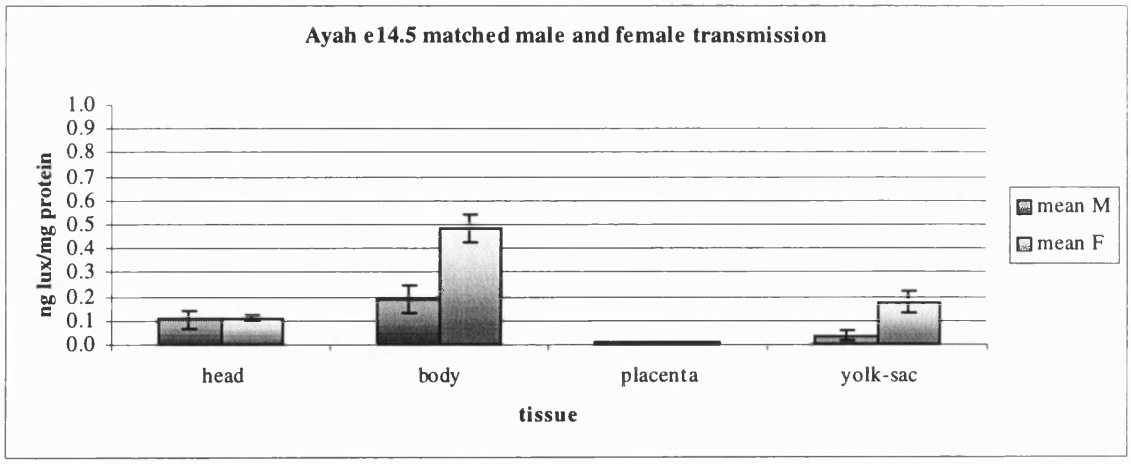


Figure 10c.

Figure 11. Parental origin-specific effects at e14.5, the ***Q*** construct.

Figure 11a. *Quark* matched male and female transmission. Mean reporter gene expression levels (ng *luciferase*/mg protein) are shown for each tissue following male (M) and female (F) transmission of the transgene. Body ($p=9.14x10^{-4}$) and yolk sac ($p=0.034$) samples show a significant difference in expression levels following maternal vs. paternal transmission of the transgene, maternal highest. Error bars show the standard error of the mean in each case.

Figure 11b. *Quasar* matched male and female transmission. Mean reporter gene expression levels (ng *luciferase*/mg protein) are shown for each tissue following male (M) and female (F) transmission of the transgene. Head ($p=3.32x10^{-4}$) and body ($p=5.79x10^{-5}$) samples show a significant difference in expression levels following maternal vs. paternal transmission of the transgene, maternal highest. Error bars show the standard error of the mean in each case.

Figure 11c. *Quiche* matched male and female transmission. Mean reporter gene expression levels (ng *luciferase*/mg protein) are shown for each tissue following male (M) and female (F) transmission of the transgene. There are no significant differences in expression between reporter genes inherited maternally compared to reporter genes inherited paternally. Error bars show the standard error of the mean in each case.

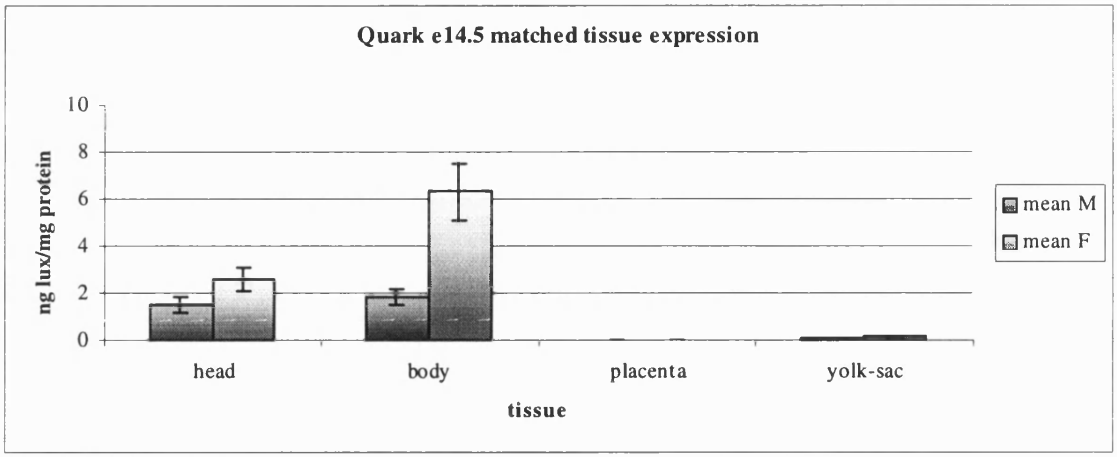


Figure 11a.

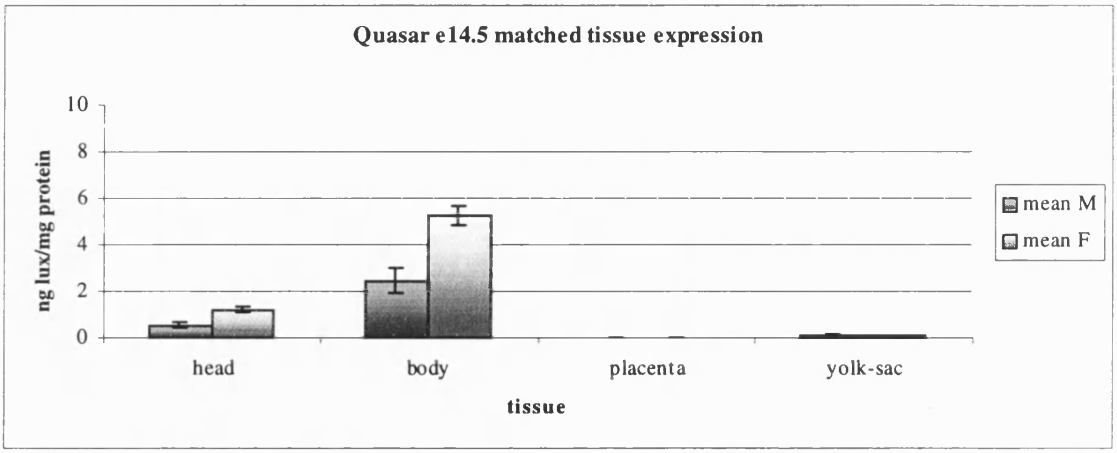


Figure 11b.

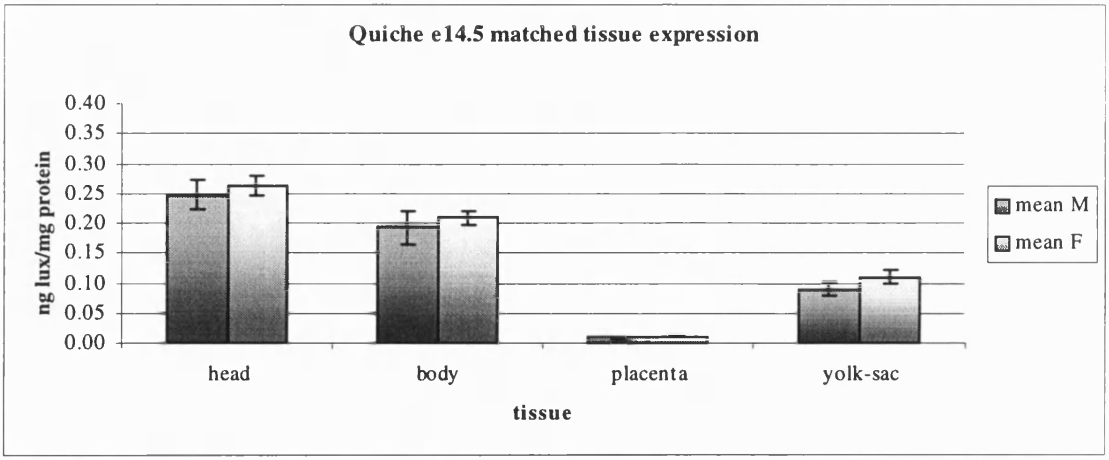


Figure 11c.

Figure 12. Parental origin-specific effects at e14.5, the ***H*** construct.

Figure 12a. *Holly* matched male and female transmission. Mean reporter gene expression levels (ng *luciferase*/mg protein) are shown for each tissue following male (M) and female (F) transmission of the transgene. Placenta samples show a significant difference in expression levels following maternal vs. paternal transmission of the transgene ($p=0.01$), maternal highest, but expression levels are generally very low. Error bars show the standard error of the mean in each case.

Figure 12b. *Hamish* matched male and female transmission. Mean reporter gene expression levels (ng *luciferase*/mg protein) are shown for each tissue following male (M) and female (F) transmission of the transgene. Head samples show a significant difference in expression levels following maternal vs. paternal transmission of the transgene ($p=9.37x10^{-4}$), maternal highest, but expression levels are generally very low. Error bars show the standard error of the mean in each case.

Figure 12c. *Harold* matched male and female transmission. Mean reporter gene expression levels (ng *luciferase*/mg protein) are shown for each tissue following male (M) and female (F) transmission of the transgene. Samples from head ($p=0.013$), body ($p=5.01x10^{-4}$) and yolk sac ($p=3.60x10^{-3}$) show significant differences in expression levels following maternal vs. paternal transmission of the transgene, but expression levels are extremely low in the latter two tissues. Error bars show the standard error of the mean in each case.

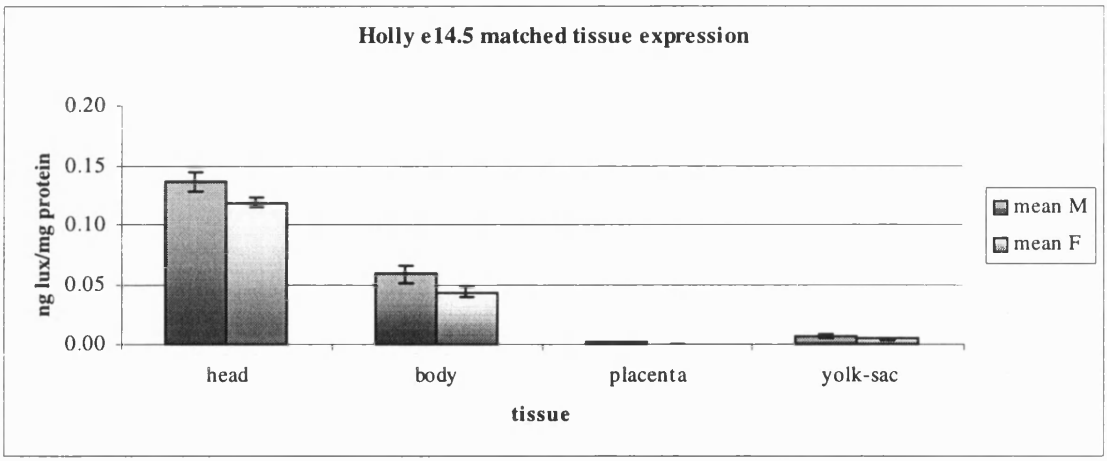


Figure 12a.

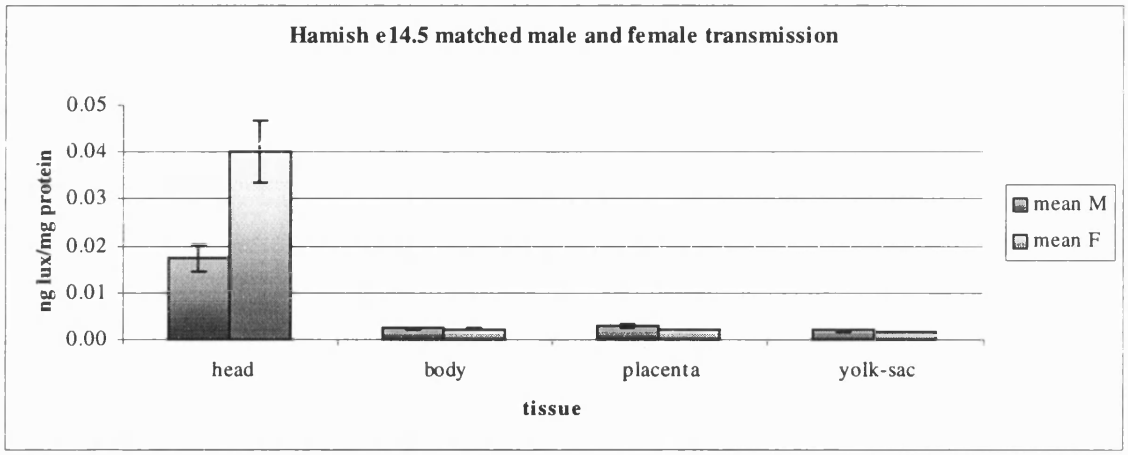


Figure 12b

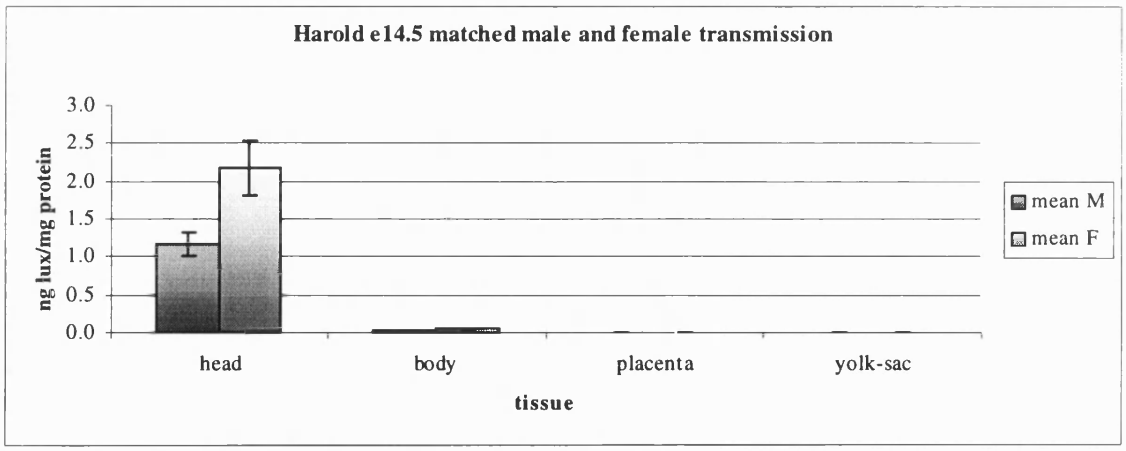


Figure 12c.

A lines (***Figure 10***)

In *Alicia* lines, a significant difference between means is seen in the body, where the mean *luciferase* specific activity following maternal transmission of the transgene is 60-fold higher than the mean specific activity following paternal transmission of the transgene ($p=7.13x10^{-4}$).

In *Axe* lines, a significant difference between the means following maternal or paternal transmission of the transgene is seen in head, body and yolk sac; in all cases the mean *luciferase* specific activity is higher following paternal transmission of the transgene. In head ($p=0.037$) and body ($p=0.006$), the difference between the parental means is less than 2-fold. In yolk sac, the ~2-fold difference between the means is at a low level of significance ($p=0.013$).

Samples from *Ayah* lines show increased *luciferase* specific activity following maternal transmission of the transgene in body and yolk sac. In the body, the level of maternally derived reporter gene expression is ~3 times that of paternally derived transgene expression ($p=6.82x10^{-4}$), in yolk sac the difference is 5-fold ($p=7.32x10^{-3}$).

Q lines (***Figure 11***)

In *Quark*, significant differences are observed between mean maternal- and paternal-allelic expression of the transgene in the body ($p=9.14x10^{-4}$) and yolk sac ($p=0.034$, only just significant). In both cases the mean *luciferase* specific activity was 3-fold higher following maternal transmission of the transgene.

Samples from *Quasar* lines show a significant difference between the means in head ($p=3.32x10^{-4}$), and body ($p=5.79x10^{-5}$). In both cases the mean *luciferase* specific activity was 2-fold higher following maternal transmission of the transgene.

No significant differences in mean reporter gene expression levels between parental samples were found in *Quiche* lines.

H lines (***Figure 12***)

Parental origin specific effects were seen in many samples from ***H*** lines where transgene expression was at basal levels, i.e., in *Harold* body ($p=5.01x10^{-4}$) and yolk sac ($p=3.60x10^{-3}$), *Hamish* yolk sac ($p=0.015$) and in the placentae of all three lines (at a low level of significance). In these cases differences between the means actually represent very small differences in gene expression, and the difference between the means was frequently less than 2-fold. Significant differences between parental means were observed in *Hamish* head

($p=9.37x10^{-4}$) where the mean *luciferase* specific activity was ~2-fold higher following maternal transmission of the transgene; and *Harold* head (at a low level of significance, $p=0.013$), where the mean *luciferase* specific activity was less than 2-fold higher following maternal transmission of the transgene.

At e14.5 parental origin-specific effects are seen in the majority of transgenic lines containing both the *H19* enhancers and the CCD. These effects in ***H*** lines are weak in the majority of cases, being present in samples where expression levels are very low (see **Chapter 4**), or at a low level of significance. The most robust parental origin specific effects (i.e. those where there is a greater than 2-fold difference between the means, at a level of significance $p<0.01$) are seen in *Alicia*, *Ayah*, *Quark* and *Quasar* lines, all of which demonstrate higher mean reporter gene activity following maternal transmission of the transgene.

Parental origin specific effects are most frequently observed in body (*Quark*, *Quasar*, *Axe*, *Alicia*, *Ayah* and *Harold*) and in yolk sac (*Quark*, *Axe*, *Ayah*, *Hamish* and *Harold*).

Table 8 summarises parental origin specific effects observed at D1 and e14.5. The rows highlighted show where D1 and e14.5 data agrees, i.e., where there is a bias in gene expression following differential parental transmission of the transgene, which is consistent in direction at both stages. *Alicia*, *Quark* and *Quasar* show consistent parental origin specific expression at the developmental stages assayed. All three lines demonstrate higher reporter gene expression levels when the transgene is maternally transmitted. ***H*** lines never show consistent parental origin specific effects, and in many cases there is either very little difference between parental means, or the differences between the means are only significant at the 5% level.

line	***Parental origin specific effects observed in:***			
	D1		**e14.5**	
	tissue	*bias*	*tissue*	*bias*
Archy	liver	paternal	nd	
	muscle			
	kidney			
Alicia	**(tongue)**	**maternal**	**body**	**maternal**
	liver			
Ayah	(brain)	paternal	body	maternal
	(liver)		yolk sac	
Axe	nd		(head)	paternal
			(body)	
			(yolk sac)	
Quark	**(brain)**	**maternal**	**body**	**maternal**
	tongue			
	liver		**yolk sac**	
Quasar	**tongue**	**maternal**	**head**	**maternal**
			body	
Quiche	(tongue)	maternal	no differences	
Holly	no differences		(placenta)	maternal
Hamish	(brain)	paternal	head	maternal
			(placenta)	paternal
			(yolk sac)	
Harold	(muscle)	paternal	(head)	maternal
			body	
	(liver)	maternal	(placenta)	
			yolk sac	paternal

Table 8. Summary of parental origin specific effects at D1 and e14.5.
Parental effects are categorised for each line according to the tissues in which they occur, and the parental bias, i.e., 'maternal' bias denotes a higher mean *luciferase* specific activity following maternal transmission of the transgene vs. the mean specific activity following paternal transmission of the transgene. 'nd' indicates analyses not done. Rows highlighted indicate those lines where a parental bias is present, and consistent at both developmental stages. Brackets indicate a low level of significance from the Student's t-test, i.e. $0.01<p<0.05$, or where there is a less than 2-fold difference between the means.

Are parental origin-specific effects a consequence of transgene methylation?

The *luciferase* specific activity assays detailed above have shown a difference in levels of transgene expression dependent upon their parent of origin. Can these differences in gene expression be explained by differences in patterns of transgene DNA methylation? CpG methylation has been associated with both differences in expression of alleles at imprinted loci (as discussed in the **Introduction**), and in silencing of multi-copy transgenes inserted into the mouse genome ([38, 39, 37, 156]).

To examine differences in methylation patterns associated with differences in reporter gene expression levels; genomic DNA samples were isolated from those individuals that show the highest or lowest levels of reporter gene activity. Samples were chosen from the livers and brains of ***A*** lines, and the livers, tongues and brains of ***Q*** lines. In liver, all three ***A*** lines, and one ***Q*** line displayed parental-allele-specific differences in gene expression. In tongue, all three ***Q*** lines showed parent-of-origin-specific effects. In brain, no such effects were seen in any line, and these samples serve as a control. DNA derived from all samples was digested with enzymes which, in combination with a specific probe, provided a diagnostic restriction pattern (see ***Figure 14***) and either *HpaII* or *MspI*. Both of these enzymes cut at the same sites (CCGG), *HpaII* is methyl-sensitive whereas *MspI* is methyl-insensitive. *HpaII/MspI* is predicted to cut the transgene once within the P3 promoter, ten times within the coding region of the *luciferase* gene, and once within the *H19* enhancers (***Figures 13*** and ***14***). Southern blots of digested genomic DNA were probed with a *P3-luciferase* specific probe, and in some cases a *H19* enhancer specific probe, which should allow the majority of the available *HpaII/MspI* sites to be tested for methylation.

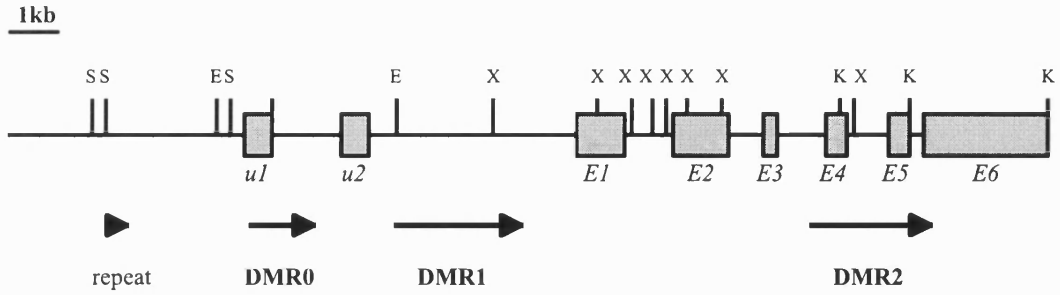


Figure 13a. 25kb genomic region including the *Igf2* gene and upstream sequences. Features include: repeat region; differentially methylated regions (DMRs) 0, 1 and 2; u*1*, *u2* are upstream exons 1 and 2. *E1-E6* are exons 1-6. Restriction sites relevent to methylation analysis are as follows: S, *SpeI*; E, *EcoRV*; K, *KpnI*; X, *XmaI*. The ***P3*** probe used for methylation analysis corresponds to 300bp immediately 5' to exon 3. Complete digestion with *XmaI/KpnI* will yield a 3.2kb fragment. Complete digestion with *SpeI/EcoRV* will yield a >15kb fragment. Both fragments will hybridise to the ***P3-luciferase*** probe.

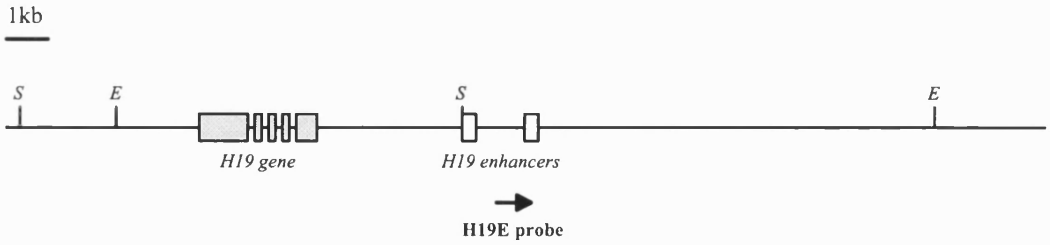


Figure 13b. 28kb genomic region encompassing the *H19* gene and the ***H19 enhancers***. Restriction sites relevent to methylation analysis are marked. S, *SpeI*; E, *EcoRV*. The ***H19 enhancer*** probe used in this analysis is indicated by an arrow below the line. Complete digestion with *SpeI/EcoRV* will yield a 12.5kb fragment that hybridises to the **H19E** probe.

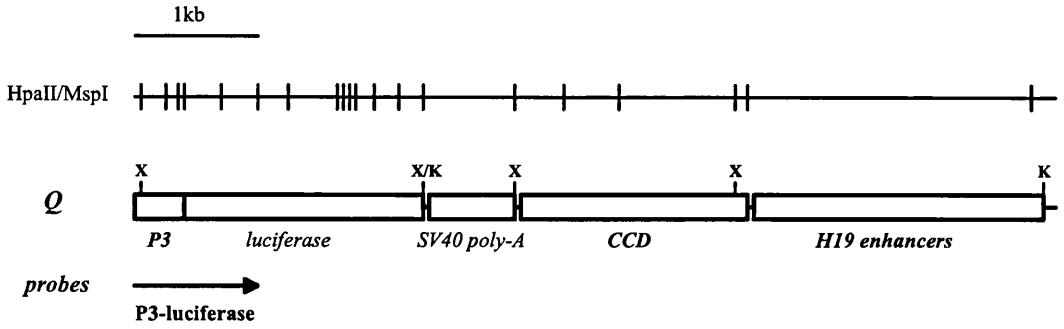


Figure 14a. The ***Q*** construct, comprised of the luciferase reporter gene with SV40 poly-A sequences driven from the *Igf2* P3 promoter, fused *in cis* to the CCD region and the *H19* enhancers. The crossed line shows *HpaII/MspI* sites within the transgene. Restriction sites relevent to methylation analysis are marked; X, *XmaI;* K, *KpnI.* The ***P3-luciferase*** probe used in this analysis is indicated as an arrow below the construct. Complete digestion with *XmaI/KpnI* should yield a 2.3kb fragment that will hybridise to the ***P3-luciferase*** probe.

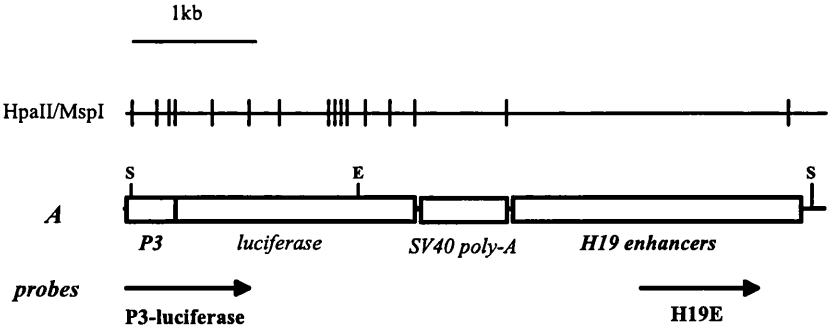


Figure 14b. The *A* construct, comprised of the luciferase reporter gene with SV40 poly-A sequences driven from the *Igf2* P3 promoter, fused *in cis* to the *H19* enhancers. The crossed line shows *HpaII/MspI* sites within the transgene. Restriction sites relevent to methylation analysis are marked;S, *SpeI*; E, *EcoRV*. The **P3-*luciferase*** probe and the ***H19 enhancer*** probe used in this analysis are indicated as arrows below the construct. Complete digestion with *SpeI/EcoRV* should yield a 1.8kb fragment that will hybridise to the **P3-luciferase** probe, and a 3.8kb fragment that will hybridise to the **H19E** probe.

Figure 15. Methylation analysis on maternally and paternally derived *Ayah* brains and livers. The origin of the samples is detailed in the table below (***Table 9***). Samples were digested with *SpeI/EcoRV* (***E/S***), *SpeI/EcoRV/HpaII* (***E/S/H***) or *SpeI/EcoRV/MspI* (***E/S/M***). Digested DNA was subjected to Southern blotting, and hybridised to probes **P3-*luciferase*** (***15a***) and ***H19E*** (***15b***). The expected fragment sizes from these digests are represented in ***Figures 13*** and ***14b***.

The transgene appears to be highly methylated in both tissues, when inherited from either parent. This methylation is present both at sites within the promoter/gene region, and at the single restriction site within the enhancer region. This methylation is not complete, as two bands at ~1.4kb and ~1kb are present in ***E/S/H*** lanes, but not in ***E/S*** lanes of ***Figure 15a***. These band appear irrespective of tissue-type or parental origin; compare brain vs. liver (***52B*** vs. ***50L***) and maternal transmission vs. paternal transmission (***42L*** vs. ***50L***).

Sample	**Parental origin of the transgene**	**Tissue**	**Specific activity** (ng *luciferase*/ mg protein)
42L	maternal	liver	0.35
43L			0.52
50L	paternal		4.97
52L			6.04
42B	maternal	brain	1.10×10^{-3}
43B			1.54×10^{-3}
50B	paternal		1.95×10^{-3}
52B			6.76×10^{-3}

Table 9. *Ayah* brain and liver samples examined for differences in DNA methylation patterns.

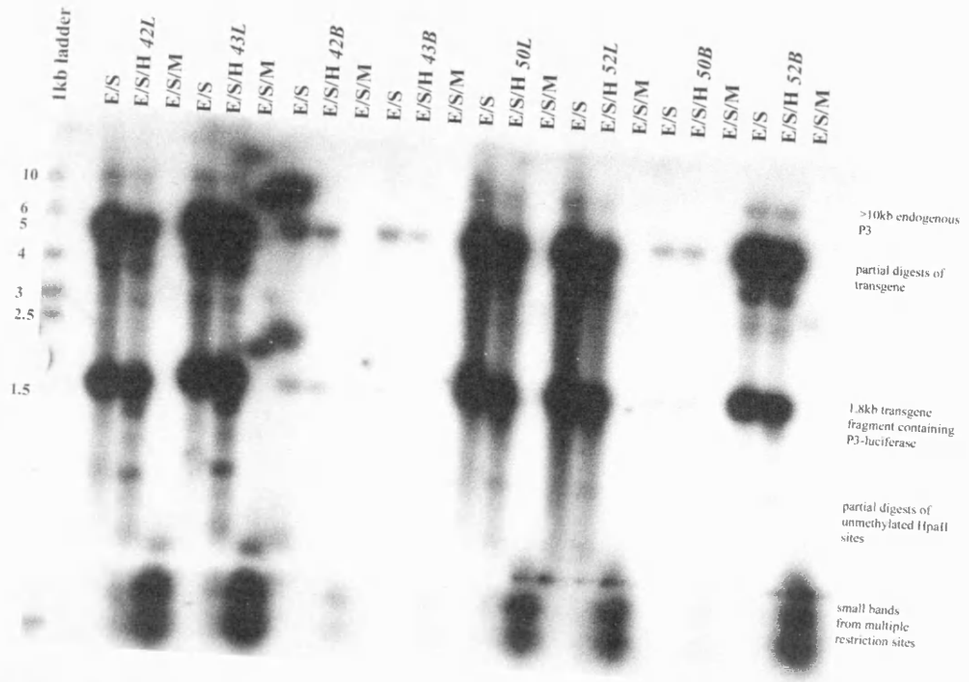


Figure 15a. P3-luciferase probe.

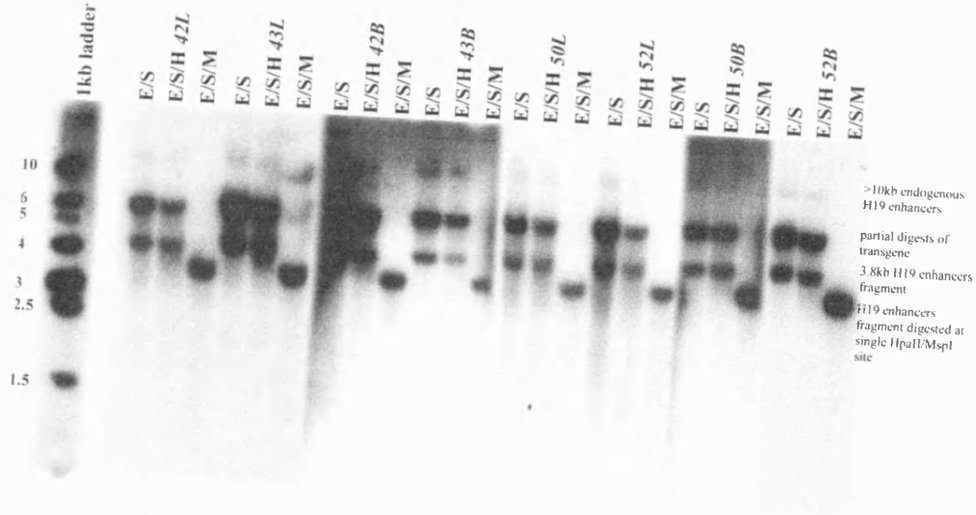


Figure 15b. H19E probe

Figure 16. Methylation analysis on maternally- and paternally-derived *Archy* brains and livers. The origin of the samples is detailed in the table below (***Table 10***). Samples were digested with *SpeI/EcoRV* (**E/S**), *SpeI/EcoRV/HpaII* (***E/S/H***) or *SpeI/EcoRV/MspI* (***E/S/M***). Digested DNA was subjected to Southern blotting, and hybridised to the **P3-*luciferase*** probe. The expected fragment sizes from these digests are represented in ***Figures 13*** and ***14b***.

The transgene appears to be under-methylated in both tissues, when inherited from either parent. Faint bands are visible at ~4kb (representing a partial transgene digest) and 1.8kb (representing a complete transgene digest) in ***E/S/H*** digests of samples ***L36*** and ***L37***, which may represent protection of the transgene by methylation in a portion of the sample. The appearance of this band does not seem to be a product of differences in loading (compare samples ***L24, B37*** and ***L36***). Samples ***L36*** and ***L37*** are both from livers where the transgene was maternally inherited.

Sample	**Parental origin of the transgene**	**Tissue**	**Specific activity** (ng luciferase/ mg protein)
22L	paternal	liver	45.07
24L			42.14
36L	maternal		2.15
37L			1.05
22B	paternal	brain	9.99×10^{-3}
24B			1.57×10^{-3}
36B	maternal		1.12×10^{-2}
37B			5.24×10^{-3}

Table 10. *Archy* brain and liver samples examined for differences in DNA methylation patterns.

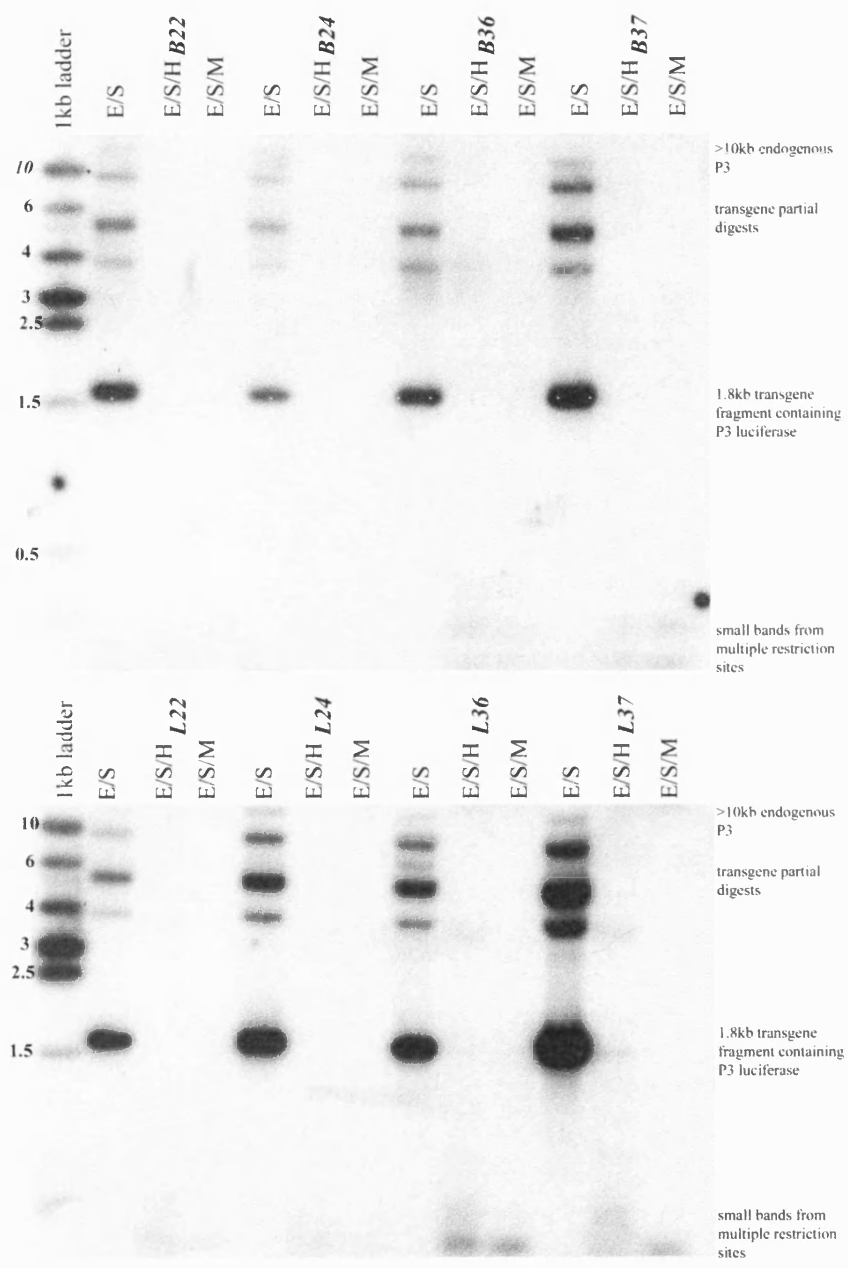


Figure 16. P3-luciferase probe.

Figure 17. Methylation analysis on maternally and paternally derived *Alicia* brains and livers. The origin of the samples is detailed in the table below (***Table 11***). Samples were digested with *SpeI/EcoRV* (***E/S***), *SpeI/EcoRV/HpaII* (***E/S/H***) or *SpeI/EcoRV/MspI* (***E/S/M***). Digested DNA was subjected to Southern blotting, and hybridised to probes P3-***luciferase*** (***17a)*** and ***H19E*** (***17b***). The expected fragment sizes from these digests are represented in ***Figures 13*** and ***14b***.

The transgene appears to be highly methylated in both tissues, when inherited from either parent. This methylation is present both at sites within the promoter/gene region, and at the single restriction site within the enhancer region. As with *Ayah* (see ***Figure 15a***), this methylation is not complete, as two bands at ~1.4kb and ~1kb are present in ***E/S/H*** lanes, but not in ***E/S*** lanes of ***Figure 17a***. These bands appear irrespective of parental origin and tissue. No differences are observed at the *H19* enhancers (see ***Figure 17b***), where the transgene is highly methylated in all samples.

Sample	**Tissue**	**Parental origin of the transgene**	**Specific activity** *(ng luciferase/mg protein)*
B12	brain	paternal	7.15×10^{-3}
B13			9.92×10^{-3}
B56		maternal	9.97×10^{-3}
B57			9.51×10^{-3}
L18	liver	paternal	6.09×10^{-2}
L19			8.81×10^{-2}
L36			1.16×10^{-2}
L37			2.83×10^{-2}
L45		maternal	27.78
L50			20.71
L59			100.30
L63			202.25

Table 11. *Alicia* brain and liver samples examined for differences in DNA methylation patterns.

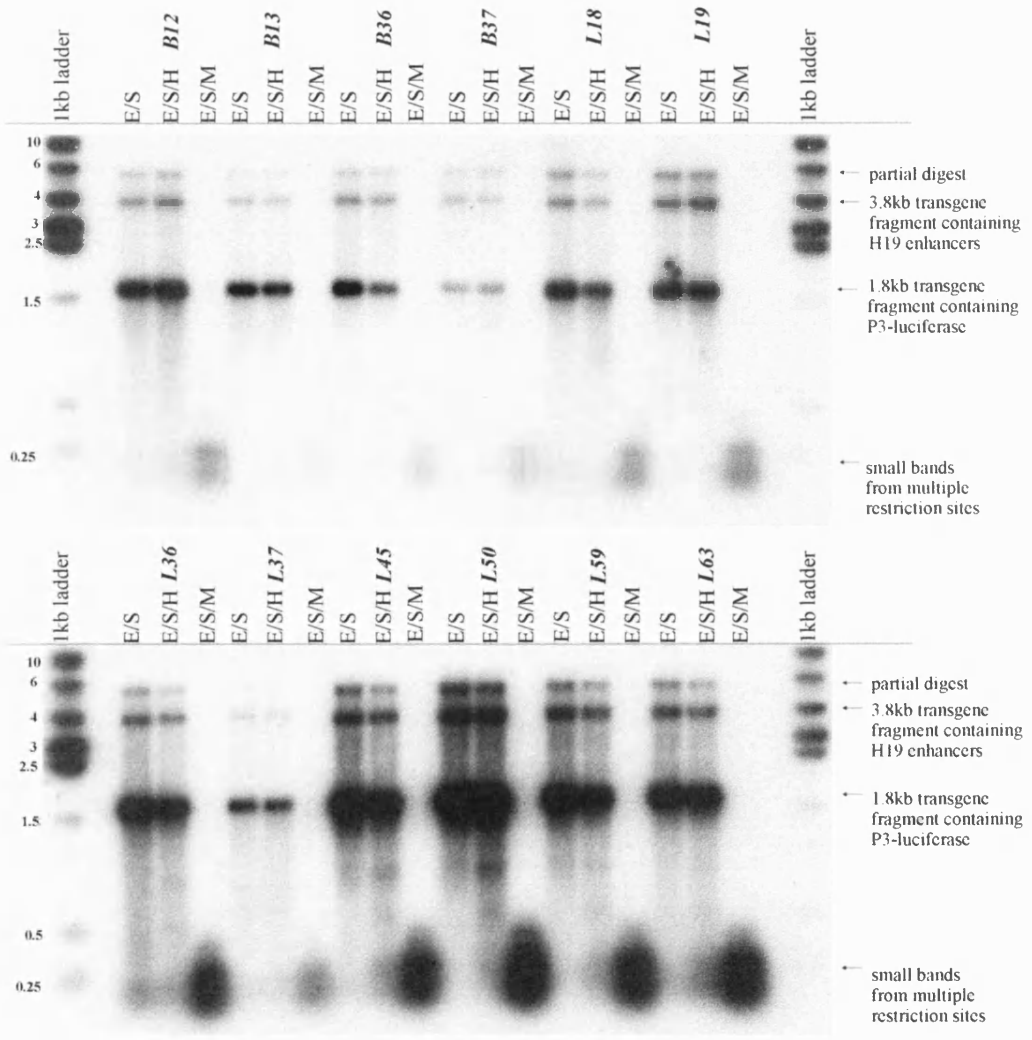


Figure 17a. P3-luciferase probe.

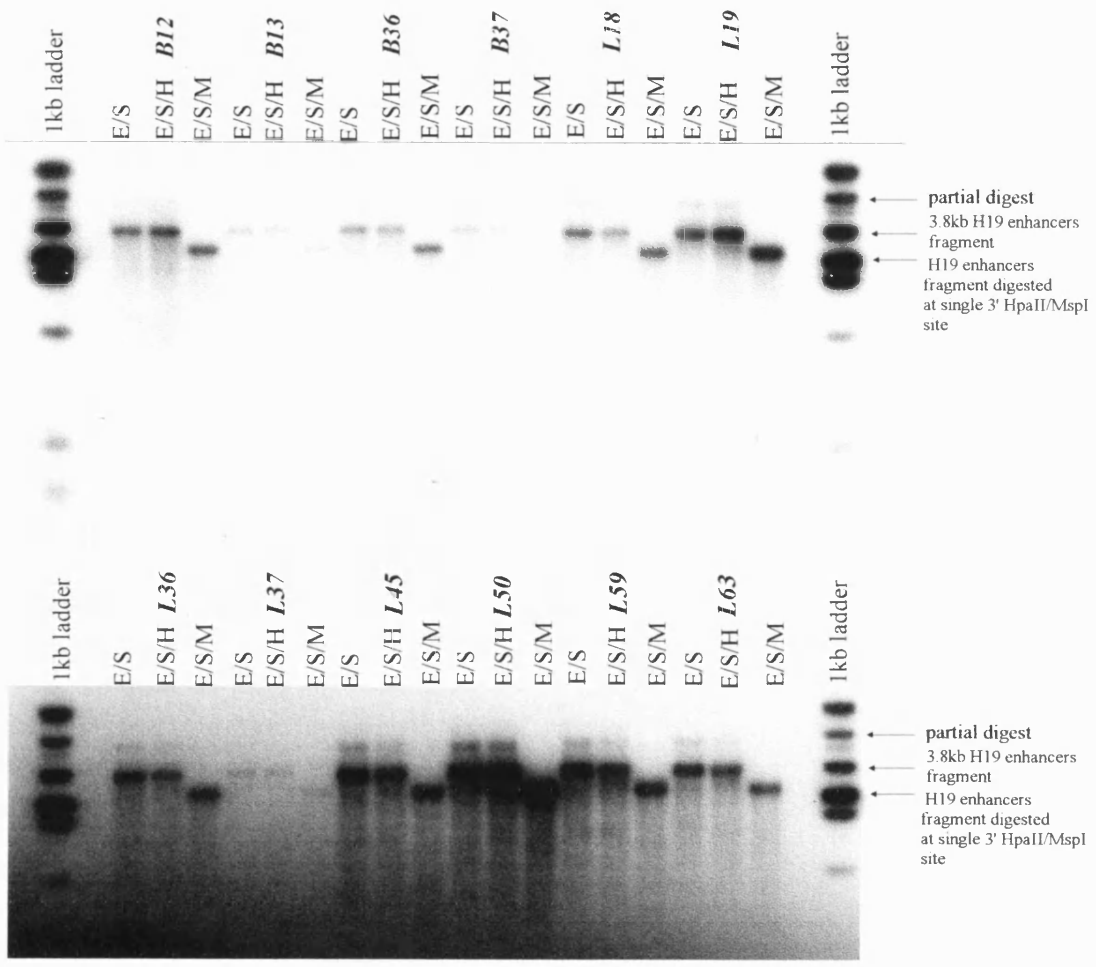


Figure 17b. H19E probe.

Figure 18. Methylation analysis on maternally- and paternally-derived *Quark* brains, tongues and livers. The origin of the samples is detailed in the table below (***Table 12***). Samples were digested with *XmaI/KpnI* (***X/K***), *XmaI/KpnI/MspI* (***X/K/M***) or *XmaI/KpnI/HpaII* (***X/K/H***). Digested DNA was subjected to Southern blotting, and hybridised to the P3-***luciferase*** probe. The expected fragment sizes from these digests are represented in ***Figures 13*** and ***14a***. ***NTG*** refers to a DNA sample derived from a non-transgenic littermate.

The transgene appears to be under-methylated in all tissues, when inherited from either parent, though unequivalent loading of brain samples complicates the analysis. The samples appear to be completely digested by *HpaII* in most cases, though in ***3L***, the 2.3kb transgene band is protected. This must be treated with caution, as no such band appears in the adjacent ***5L*** samples, where the DNA loading is of equivalent quantity.

Sample	**Parental origin of the transgene**	**Tissue**	**Specific activity (ng *luciferase*/ mg protein)**
3B	paternal	brain	2.70×10^{-2}
39B	maternal		1.73×10^{-2}
40B			3.00×10^{-2}
3L	paternal	liver	1.73
4L			2.03
5L			1.49
39L	maternal		9.19
40L			10.75
42L			9.81
2T	paternal	tongue	1.67×10^{-4}
3T			1.52×10^{-4}
4T			1.22×10^{-4}
39T	maternal		9.04×10^{-4}
40T			8.06×10^{-4}
41T			6.96×10^{-4}

Table 12. *Quark* brain, tongue and liver samples examined for differences in DNA methylation patterns.

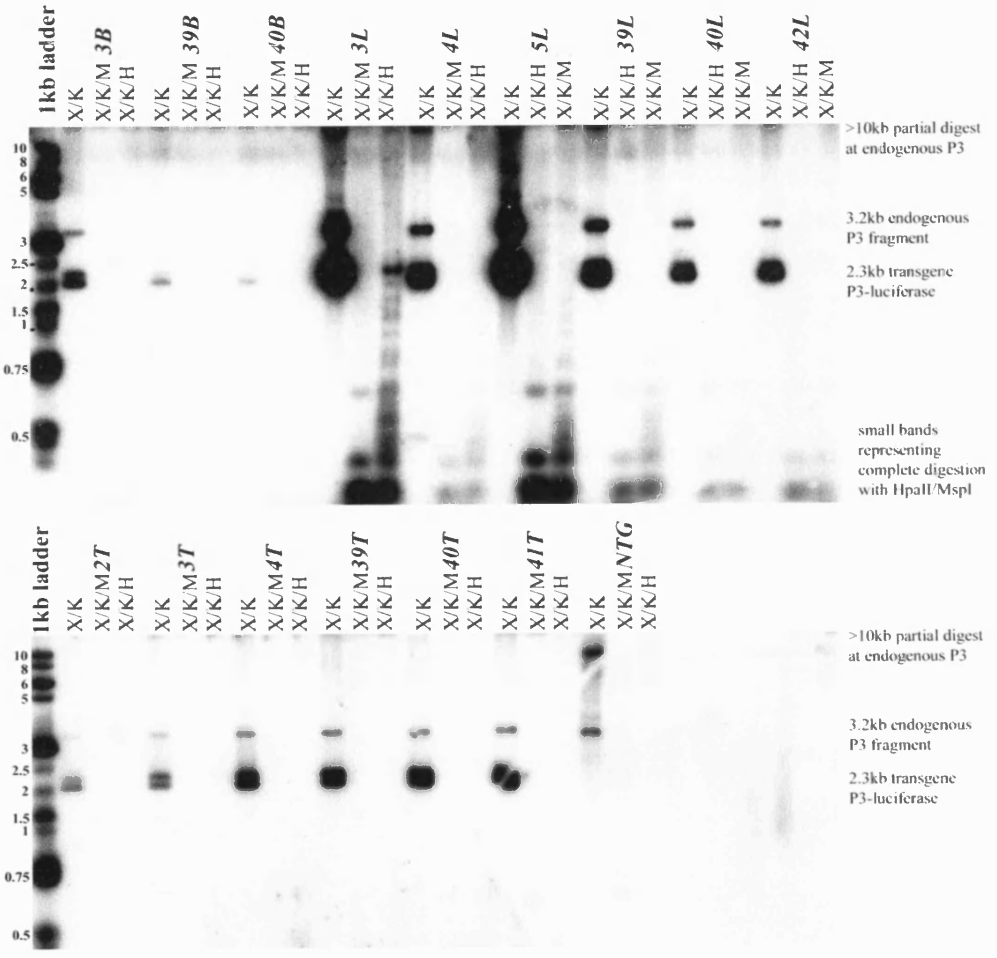


Figure 18. P3-luciferase probe.

Figure 19. Methylation analysis on maternally and paternally derived *Quasar* brains, tongues and livers. The origin of the samples is detailed in the table below (***Table 13***). Samples were digested with *XmaI/KpnI* (***X/K***), *XmaI/KpnI/MspI* (***X/K/M***) or *XmaI/KpnI/HpaII* (***X/K/H***). Digested DNA was subjected to Southern blotting, and hybridised to the **P3-*luciferase*** probe. The expected fragment sizes from these digests are represented in ***Figures 13*** and ***14a***. ***NTG*** refers to DNA samples derived from non-transgenic littermates.

The transgene appears to be under-methylated in all tissues, when inherited from either parent.

Sample	**Parental origin of the transgene**	**Tissue**	**Specific activity** (ng *luciferase*/ mg protein)
13B	maternal	brain	0.75
23B	paternal		0.68
53L	paternal	liver	6.11
64L			7.92
65L			6.97
13L	maternal		8.70
49L			7.84
73L			7.96
53T	paternal	tongue	2.34×10^{-3}
55T			3.63×10^{-3}
56T			6.42×10^{-3}
12T	maternal		0.15
13T			0.20
15T			0.17

Table 13. *Quasar* brain, tongue and liver samples examined for differences in DNA methylation patterns.

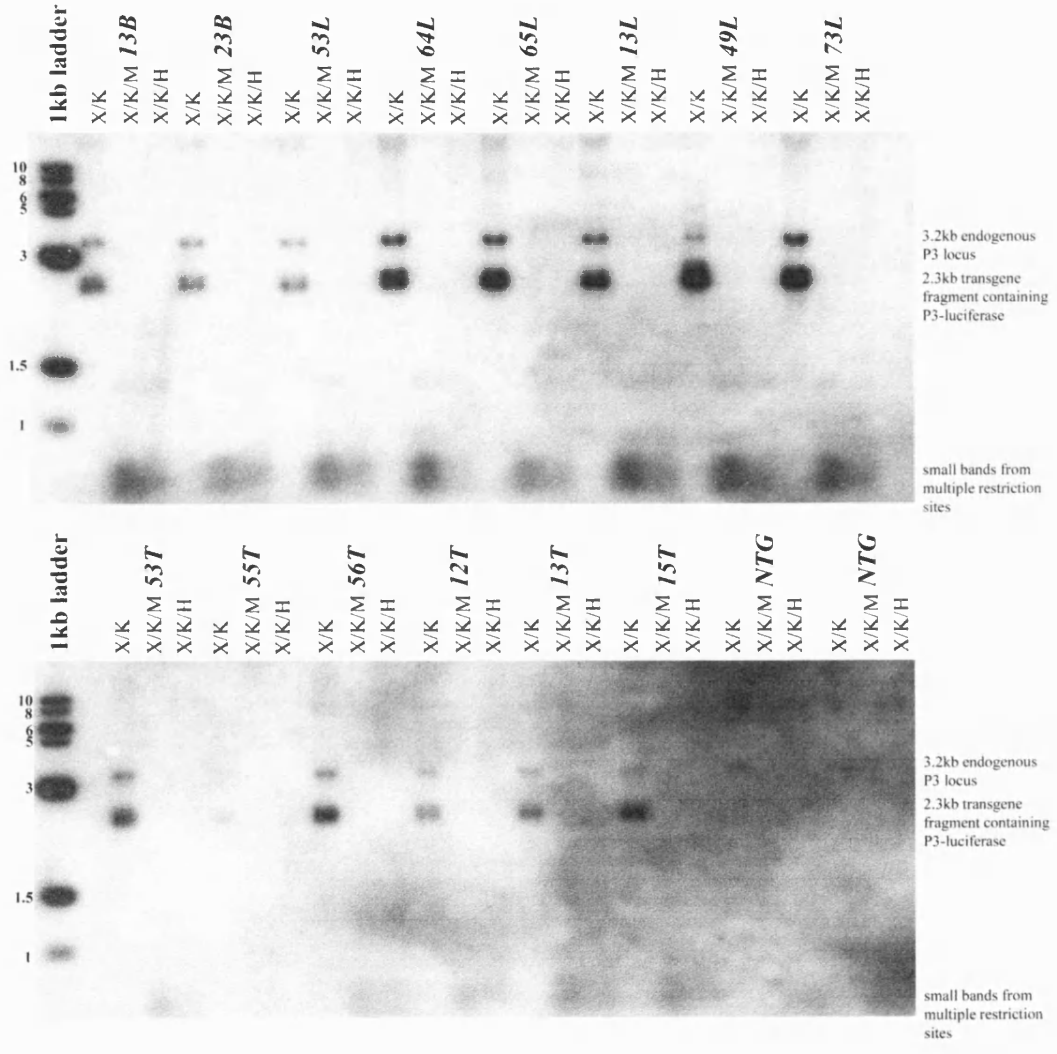


Figure 19. P3-luciferase probe.

Summary of methylation analysis

A lines

In *Ayah* lines (***Figure 15***), DNA from brain and liver exhibits almost complete methylation of the transgene, both at P3-*luciferase* and at a single site at the *H19* enhancers. Partial methylation of the transgene is exhibited by fragments of 1.4kb and 1kb that hybridise to the P3-*luciferase* probe. The *HpaII* sites giving rise to these fragments must be within the coding region of the *luciferase* gene, as opposed to the promoter, as the promoter region is only ~0.3kb long, giving a minimum fragment size of 1.5kb (see ***Figure 14b***). Despite differences in DNA loading the methylation patterns of the transgene appear to be identical in brain and liver, following male and female transmission of the transgene.

In *Archy* lines (***Figure 16***) the ***A*** transgene appears to be undermethylated in both brain and liver. Faint bands corresponding to the full length 1.8kb transgene are present in liver samples derived from maternal transmission of the transgene, which are known to display *luciferase* specific activities 20-40-fold lower than their paternally derived counterparts (***Table 10***). The differences between mean *luciferase* specific activity following male vs. female transmission of the transgene in *Archy* liver is highly significant ($p=3.70x10^{-8}$). These partial digests are absent from other brain and liver samples, and represent a minority of the total transgene DNA.

In *Alicia* lines (***Figure 17***), the transgene exhibits almost complete methylation, both at P3-*luciferase* and at the *H19* enhancers. Partial methylation of the transgene is identical to that exhibited by the ***A*** transgene in *Archy* lines.

Q lines

Quark samples from liver, brain and tongue generally exhibited a pattern of methylation where the transgene was entirely unmethylated (***Figure 18***). Unequivalent loading prevents detailed examination of brain samples, but they appear to be largely unmethylated. An exception lies in a single liver sample, where a full-sized transgene fragment is present in *HpaII*-digested DNA, representing protection of transgene digestion by methylation. This band is unique amongst the liver samples, so hypermethylation is not taken to be a general feature of the transgene in *Quark* liver.

The transgene methylation patterns observed in *Quasar* livers, brains and tongues are identical to those of *Quark*, i.e., the transgene was found to be fully unmethylated in all tissues, regardless of the parental origin of the transgene (***Figure 19***).

To summarise, samples with *luciferase* specific activities ranging from ~$1x10^{-4}$ to ~$1x10^{2}$ were assayed for levels of transgene methylation, and no correlation was found between reporter gene activity and transgene methylation. The methylation status seemed to be reliant upon the particular line examined (i.e., in the *Quasar* line the transgene is unmethylated, in the *Alicia* line the transgene is almost fully methylated), and all samples within, regardless of tissue or parental origin of the transgene, displayed the same methylation pattern.

Conclusions

Can either the *H19* enhancers or the CCD confer parental origin-specific differences in expression on the reporter gene? As seen in a previous study [108], the *H19* enhancers reproducibly confer parental-origin-specific differences in *luciferase* expression from P3. In the experiments described above, at D1 all transgenic lines studied containing this element demonstrated significantly higher *luciferase* specific activities when the transgene was inherited either maternally (*Quark, Quasar, Quiche, Alicia*), or paternally (*Archy, Ayah*). In one case, the ratio of parental transmission was as high as 14:1 (*Quasar* tongue). These imprinting-like effects were most commonly observed in liver and tongue, and interestingly, were never observed to a high level of significance in brain, where *Igf2* is known to be expressed biallelically [70]. The maternal allele of *Igf2* must bypass the imprinting mechanism in the choroid plexus and leptomeninges of the brain, either by escaping inactivation during early development, or by being reactivated at a later developmental stage. Support for the second of these mechanisms comes from a study of *Igf2* expression in the rat choroid plexus [102], where monoallelic expression from the paternal allele is seen on days 13.5 and 15.5 of gestation, and a subsequent switch to biallelic expression is observed by e18.5. Thereafter approximately equivalent expression from both alleles is observed until at least 3 months of age. The absence of parental allele-specific effects in brains in this study may be due to an absence in these tissues of a *trans*-'imprinting factor' required to silence the maternal allele. The presence of parental origin-specific effects in *Q*-lines demonstrates that the CCD is not able to override the effect of the *H19* enhancers.

At e14.5, parental origin-specific effects were again observed in the majority of lines, but the effects were most robust in those lines that contained the *H19* enhancers (*Alicia, Ayah, Quark and Quasar*). In all cases these lines demonstrated greater reporter

gene activity following maternal transmission of the transgene. Consistent parental origin-specific effects were never observed in ***H*** lines, and in many cases there was either very little difference between parental means, or the differences between the means only were significant at the 5% level.

The investigation of the tissue-specificity of these effects was limited, as the dissections were very crude. Despite this, some tissue specificity was apparent as if parental origin specific effects were observed; they were always manifested in the bodies of transgenic embryos, and often observed in the yolk sac (*Ayah, Quark* and *Quasar*). None of the transgenic lines tested expressed *luciferase* above background levels in the placenta, so no genuine parental origin specific effects could be observed in these samples.

A comparison of the data from the D1 study and the e14.5 study reveals that *Alicia, Quark* and *Quasar* lines consistently demonstrated a higher level of *luciferase* specific activity following maternal transmission of the transgene. The magnitude of the expression level following maternal vs. paternal transmission ranged from ~2-fold (*Quasar*) to 60-fold (*Alicia,* e14.5). A maternal bias conferred on *luciferase* gene activity by the *H19* enhancers was also reported in a previous analysis of some of the transgenic lines discussed here [108]. It cannot be concluded, however, that the only 'real' parental origin-specific effects were those where expression levels were highest following maternal transmission of the transgene, as one of the most significant effects was observed in *Archy* liver (p=3.70x10^{-8}), where expression levels were ~4-fold higher following paternal transmission of the transgene. In this case, the comparison at e14.5 was not performed.

It is difficult to compare these results to other studies in transgene imprinting, as in most cases the levels of expression from individuals bearing transgenes following maternal or paternal inheritance are not quantitated [38, 39, 37]. In addition the variation in expression levels between individuals of a particular parental cross are not reported. In this study, large differences in gene expression were often observed between individuals that had inherited the transgene from one-or-other parent (e.g. see ***Table 13***), which were reminiscent of the 'high/low' levels of expression observed at endogenous imprinted loci. However, the differences in reporter gene specific activity were often observed against the background of individuals expressing the reporter gene at intermediate levels. In one study [109] transgenic mice were created with constructs bearing *LacZ* or *placental alkaline-phosphatase (PLAP)* reporter genes *in-cis* to the *H19* enhancers and a portion of the *H19* upstream region. When the transgene construct contained at least 3.8kb of the *H19* 5'

region, the transgenic embryos demonstrated repression of the reporter gene upon paternal transmission, and activation upon maternal transmission. The degree of repression upon paternal inheritance was variable between littermates, and in a minority of cases the embryos escaped inactivation altogether. At the level of the tissue, partial imprinting effects can be observed, with genomic elements biasing the contribution of gene expression from one-or-other of the parental alleles. It would be interesting to discover whether expression differences between parental alleles in a given tissue are due to changes in quantitative expression levels in each cell, or a change in the number of cells expressing the gene.

Expression from the maternal allele of the *Igf2* gene has been reported to occur in 8.3% of hepatocytes at e13.5 [111], and maternal *Igf2* mRNA is estimated at 2-4% of paternal levels in MatDi7 and parthenogenetic embryos [8]. Similarly, paternal *H19* transcripts have been estimated to represent up to 5% of late-embryonic and neonatal *H19* mRNA [86, 27], and expression from this allele was reported in 13.4% of hepatocytes at e13.5 [111]. Ratios of mean reporter gene expression following differential parental inheritance of 14:1 (*Quark* tongue at D1) and 60:1 (*Alicia* body at e14.5) compares favourably with the reports of expression from the 'silent' allele of imprinted genes at a level of ~5-10%. An assumption of this comparison is that *luciferase* specific activity can be taken as a direct indicator of gene expression, (i.e., that the amount of *luciferase* protein measured in the above assays is directly proportional to the level of *luciferase* mRNA), which may not be the case. The majority of parental origin-specific effects reported here show a 2-3-fold difference in mean *luciferase* specific activity following differential parental transmission of the transgene, in line with a previous study of these transgenes [108]. Taken together, these results lead to the conclusion that the *H19* enhancers are playing a minor role in biasing gene expression levels in a parent of origin-specific manner.

The link between cytosine methylation and imprinting is well documented, and discussed at length in the **Introduction**. Studies of transgenes constructed from parts of the *H19* locus [38, 39, 37] have reported hypermethylation of the silenced transgene both at the promoter and within the structural gene (which are differentially methylated at their endogenous location). A *H19* gene deletion was constructed [160], in which almost the entire transcription unit was replaced with the *luciferase* gene. When this modified *H19* locus is maternally inherited *luciferase* is expressed to high levels in liver and skeletal muscle. There is an appropriate lack of methylation at the *H19* promoter, and of the

upstream region on the maternal allele. When paternally inherited the *luciferase* gene shows a large variability in expression levels, from 5-55% the level of maternal *luciferase* expression. The level of *luciferase* expression from the paternal allele is inversely correlated to the degree of methylation of the promoter and the gene. An individual expressing the paternal allele at 9% of the maternal expression level displayed low levels of gene methylation. In an individual expressing the paternal allele at 24% of maternal levels, the fractions of methylated to unmethylated DNA were approximately equal. This study shows that the *luciferase* gene can be silenced by methylation, at least when driven from the *H19* promoter, and that detectable differences in gene methylation can be correlated to expression levels.

With this in mind, it seemed reasonable to ask whether the parental-origin specific differences in gene expression observed in *H19* enhancer-bearing P3-*luciferase* constructs could be correlated to differences in transgene methylation following maternal or paternal transmission. It was found that differences in the methylation status of the transgene following differential parental transmission generally do not appear to exist, even where *luciferase* gene expression levels between individuals vary by ~2000-fold (*Alicia* liver, samples ***L36*** vs. ***L63, Figure 17***). One exception to this is in *Archy* liver, where samples are generally undermethylated; in samples deriving the transgene maternally, a faint band is present in the *HpaII*-digested lane (***Figure 16***), corresponding to a small proportion of completely methylated transgenes. It is worth noting that *Archy* livers displayed the most significant difference between mean reporter gene activities following maternal vs. paternal transmission of the transgene. Despite this, gross changes in DNA methylation at the transgene do not reproducibly correspond to differences in reporter gene expression level. It is possible that these experiments lack the sensitivity required to detect the changes in methylation status responsible for the gene expression differences. As all the lines examined above may contain the transgenes in multi-copy arrays (see ***Appendix 1***), the analysis is complicated further, as e.g. demethylation of a single site within the promoter of a highly methylated transgene array may be swamped by the hypermethylated signal. In addition, only a single potentially methylated restriction site was examined in the *H19* enhancer region of the transgene. If transgene methylation is involved in the differential silencing of this locus, the pattern of methylation does not resemble that of the *luciferase* gene 'knock-in' study cited above [160], where expression levels correlated to large changes in CpG methylation over several sites, including sites within the *luciferase*

reporter gene sequences. Instead, silencing of P3-*luciferase* reporter constructs would have to be reliant on the methylation status of very few crucial CpG residues.

A surprising result of the methylation analysis was that samples in which the *luciferase* specific activity measurements were very high could still be very highly methylated. Some studies of methylation at transgenic loci [209, 210, 211] report a strong correlation between transgene expression levels and methylation. An *RSVmyc* transgene was found to be expressed, when paternally inherited, in the myocardium of the heart [209], and silent when maternally inherited. The active expression state corresponded to a hypomethylated (but not unmethylated) transgene, and the silent state to a hypermethylated transgene, at several different genomic insertion sites [211]. Conversely, a group studying expression of a *CAT* reporter gene from the promoter of the *HMGCR* (3-hydroxy-3-methylglutarylCoA-reductase) gene found no correlation between *CAT* gene expression and methylation of the transgene [208]. At its endogenous locus, the *Igf2* gene has been found to be highly methylated both in the upstream region [128] and in the body of the gene [8, 7] on the expressed paternal allele.

Genetic background effects have been shown to play a strong role in the expression and methylation status of transgenes [210, 211]. A *HSVtk*-promoter-*LacZ* fusion gene was used to create a transgenic founder on a complex genetic background [210]. When this founder was backcrossed onto a parental strain, the F1 generation displayed a characteristic bimodal pattern of gene expression, in which the foetuses expressed *β-galactosidase* at high, or low levels, in equal proportions. The F2 generation displayed four expression level phenotypes, from 'very low' to 'very high'. *LacZ* expression varied in every cell more or less equally, and with little cell to cell variation in the intensity of *β-galactosidase* staining, so the differences in individual expression profiles were not due to position-effect-variegation. The same transgene was bred onto a variety of pure-breeding genetic backgrounds, many of which affected the expression of the transgene following a Mendelian pattern of inheritance. The investigators concluded that unlinked modifier genes were present in some mouse strains which could effect the expression (and methylation status) of the transgene. The *RSVmyc* transgene mentioned above also is sensitive to genetic background effects [211]. *RSVmyc* transgenes are expressed following paternal transmission in the heart, and silent following maternal transmission, (as discussed above)

in the original complex founder strain, and on an inbred FVB/N background. However, on a C57BL/6J background, the transgene is always silent.

The transgenic parents used to generate offspring whose tissues were analysed in this study were the result of at least 3 generations of intercrosses of (C57Bl/6xCBA) F1 mice, and so reflect a complex genetic background. Though male and female sibs were chosen to normalise for parental age and generation differences, the parents nor the offspring would be predicted to be genetically identical. The gene expression differences apparently due to parental origin of the transgene could therefore be due to the action of modifier genes present in some parents, but not others, and could therefore be independent of parental sex. This argument is weakened by the fact that more than one parent of a particular sex was used to generate offspring, at two different developmental stages. However, the action of modifiers can only be ruled out by examination of progeny that are genetically homogeneous, differing only in the parental origin of the transgene.

A tissue-specific, parental allele-specific effect has been demonstrated with P3-*luciferase* transgenic lines that contain the *H19* enhancers. *Alicia, Quark* and *Quasar* lines all demonstrate higher levels of reporter gene expression following maternal expression of the transgene, at both D1 and at e14.5. Addition of the CCD into reporter constructs does not prevent these parental origin specific effects, unless it does so in combination with other elements not examined by this study, or in further tissues that were not studied. These parental-origin-specific differences in gene expression are not correlated to gross changes in the level of transgene methylation.

CHAPTER 6, IS THE CCD ABLE TO DIRECT REPORTER GENE EXPRESSION IN AN *IN-VITRO* SYSTEM? PART 1, IMMORTAL CELL LINES.

Introduction

In parallel to the study of expression of P3-*luciferase* reporter constructs *in vivo* (see **Chapter 4**), expression of the reporter constructs (***Figure 3*, Chapter 1**) in tissue specific cell lines was examined by transient transfection experiments.

Transient transfection systems provide a sensitive method by which to examine and delineate potential tissue-specific regulatory elements. Transient transfection of the hepatocarcinoma cell line *Hep3B* with DNA encoding the *H19* gene and surrounding regions led to the localisation and characterisation of the *H19* enhancers [103]. By the creation of a series of overlapping subfragments of an original 22.5kb parent construct (containing the *H19* gene, 11kb of the 5' flanking region and 8kb of the 3' flanking region), liver-specific enhancer activity was localised initially to a 2.4kb region approximately 10kb 3' to the start site of the *H19* gene. By more detailed mapping, this activity was confined to two separable elements, each approximately 270-300bp in length, which were each able to enhance transcription of the *H19* gene ~30-fold in hepatoma cells. Together these elements acted additively to give a ~70-fold enhancement of gene expression from the *H19* promoter in an orientation-independent manner.

The main question posed in the following experiments is; can the CCD drive reporter gene expression in and *in vitro* cell culture system? The eventual aim being to isolate subregions within this domain that contain regulatory activity. A necessary preliminary to this investigation was to determine if P3-*luciferase* reporter constructs demonstrated the same tissue specific gene activity as described for the *H19* reporter constructs described above. To this end, *Hep3B* cells were transiently transfected with the ***M*** construct, and the ***A*** construct (***Figure 3***) in order to determine whether *luciferase* gene expression driven from the P3 promoter was upregulated when the *H19* enhancers were *in cis*.

The *H19* enhancers provide the majority, if not the entirety of liver expression of *Igf2* and *H19* [104], so it is unlikely that the CCD acts as an enhancer in this tissue, and therefore in *Hep3* cells. The ***H*** construct (***Figure 3***) was transiently transfected into these

cells to provide confirmation of this, and to test the possibility that the CCD acts in a tissue-independent manner. The CCD may however, act to modulate activity from the known enhancers, rather than acting autonomously. The ***Q*** construct (***Figure 6***) was transfected into this cell line to test this possibility.

The *C2* cell line, originally isolated from murine primary myoblasts [187] provides a model system in which to study the expression of *Igf2* in muscle development.

Igf2 has been shown to be expressed in the mesoderm of the embryo from the earliest stages of its development [87]. *Igf2* expression persists to high levels in all three regions of the somites (sclerotome, dermatome, and myotome) and in their derivatives in later development. The expression in myotome increases as development proceeds, and the levels of *Igf2* remains high in myoblasts as they differentiate into myotubules [95].

C2 cells can be induced to differentiate into myotubules upon serum withdrawal, or addition of Horse Serum [187, 188]. This change is detectable both by morphological changes in the cells, and the expression of muscle-specific proteins. Among these proteins are many of the factors required for *IGF* signalling, including *Igf1, Igf1R* and an *Igf binding protein* [212], as well as *Igf2*.

Igf2 is expressed at low levels in undifferentiated *C2* cells, as might be expected if these cells resemble native myoblasts [95, 213]. Upon induction of differentiation of *C2* cells, *Igf2* mRNA levels begin to rise; after 16 hours, a 2-fold upregulation over basal levels can be detected, this level rises to reach a 15-fold upregulation by 48 hours, and a peak of a 25-fold upregulation by 96 hours. The upregulation of *Igf2* expression roughly follows the morphological changes that occur in the cells, as by 48 hours following induction of differentiation, cell fusion and the formation of myotubules is well underway. By 96 hours, giant myotubules have developed [212].

The transcription of the *Igf2* gene in embryonic tissues has been found to be initiated predominantly from P3 [214], though a very small percentage of transcripts are synthesised from promoter 2. A *luciferase* reporter gene driven from P3 has been shown to be expressed at high levels when transiently transfected into undifferentiated *C2* cells, but the promoter sequences alone were not sufficient to mediate the upregulation of this gene upon differentiation by serum withdrawal. Investigation of 25kb of the *Igf2* proximal region by this method did not isolate any sequences which were able to enhance gene expression *in cis* to P3 upon myoblast differentiation.

The *H19* enhancers have been reported to drive the expression of reporter genes in some mesodermal lineages (sclerotome, tongue muscle [109, 106]). However, it is doubtful that these enhancers drive skeletal muscle-specific expression of *Igf2* or *H19*, as a germ line deletion of the region containing the enhancers did not significantly reduce the expression of either of these genes in this tissue [104].

The CCD provides a candidate region for a mesoderm-specific enhancer of *Igf2* and/or *H19*. If this element does drive expression in mesodermal tissues, specifically in skeletal muscle, the ***H*** reporter constructs might be expected to express *luciferase* when transiently transfected into the *C2* cell line, and to be upregulated upon their differentiation into myotubules.

Results

Construct DNA (depicted in ***Figure 3***) was transiently transfected into a number of cell lines, to allow examination of *luciferase* activity when P3 is *in cis* to the *H19* enhancers (***A*** construct), the CCD (the ***H*** construct), and both elements together (the ***Q*** construct). Levels of expression from P3 alone (***M*** construct) was measured in all experiments, and provided a baseline to which other construct activity was compared. In all cases a construct containing the *LacZ* gene driven from the *SV40* promoter was co-transfected with test constructs to control for transfection efficiency, and levels of total protein were measured to control for cell number. All figures of reporter construct activity are calculated as ng *luciferase*/mg total protein/level of *β-galactosidase* activity. When no *luciferase* constructs were present, expression was never detected over background levels for the *luciferase* assay.

Expression levels from test constructs was compared to expression from the ***M*** construct by the non-parametric Wilcoxon's matched pairs test. As sample sizes were small a normal distribution of the data could not be assumed. A paired test was used due to variation in basal activity between experiments. The null hypothesis in each case was; 'There is no difference in reporter gene expression following transient transfection of the basal construct vs. the test construct'. All tests were conducted at a 5% significance level.

Hep3B cells

Luciferase expression levels following transient transfection of construct DNA into *Hep3B* cells are tabulated below, and ratios of test construct to basal level are noted (***Table 14***).

The ratios of mean test construct expression over basal expression are also presented in graphical form (***Figure 20***).

Following transient transfection into *Hep3B* cells the ***H*** construct exhibited no significant upregulation of reporter gene activity over basal levels. Conversely, the ***A*** construct enhanced transcription of *luciferase* from P3 by approximately 20-fold, and the ***Q*** construct enhanced transcription of *luciferase* by approximately 10-fold. There is a significant difference between the expression levels from ***A*** and ***Q*** constructs at the 5% level, suggesting that the CCD is reducing expression from the *H19* enhancers by a factor of ~2-fold.

C2 cells

Following transient transfection into undifferentiated *C2* cells, none of the test constructs exhibited increased reporter gene expression over basal levels, see ***Table 15*** and ***Figure 21***.

Transfected *C2* cells were induced to differentiation by the addition of horse serum, and expression from reporter constructs was measured 48 hours following this induction. At 48 hours post-induction the cells had visibly differentiated, as judged by the observation of myotubules.

The expression of reporter constructs in differentiated *C2* cells does not appear to differ greatly in magnitude between test constructs and basal levels. Test constructs express the reporter gene at between 50% and 75% the levels of basal expression see ***Table 16*** and ***Figure 22***. The sample size was too small ($n=4$) to determine whether the differences between basal levels and test construct levels of expression were significant. Instead, a Mann-Whitney-U test was performed to compare expression level differences between differentiated and undifferentiated cells. Comparisons were made between the expression levels from the ***M*** construct when the cells were undifferentiated vs. differentiated ($n=12$). This test was repeated with the ***A*** construct and with the ***Q*** construct. In each case, no significant differences were found in expression of the *luciferase* reporter gene between differentiated and undifferentiated *C2* cells, when transfected with a particular construct.

Further cell lines

Preliminary experiments were performed on two further cell lines, *Cos7* cells and *NIH 3T3* cells.

In each cell line no enhancement over basal levels of expression were apparent with any test construct (data not shown). Sample sizes were small in both cases, so these results remain speculative, though it seemed clear that in no case was reporter gene activity greatly enhanced over basal levels of expression.

Table 14. Summary of transient transfection data for *Hep3B* cells. The ***M*** construct is treated as providing the basal level of expression in this experiment, to which all other constructs are compared. A normalised value of test construct expression is presented as the ratio of test construct expression vs. basal (***M***) expression. The mean, variance and standard error of the mean (s.e.) were calculated for each construct, and each normalised construct, where $n=7$.
There was a significant difference at the 5% level between ***M*** and ***A***, between ***M*** and ***Q***, and between ***A*** and ***Q***.

Figure 20. Expression levels of each test construct (***H, A*** and ***Q***) following transient transfection into *Hep3B* cells. All test constructs was normalised by dividing by the level of basal expression, which is judged as expression levels following transfection of the ***M*** construct (see ***Table 14***). Error bars show the standard error of the mean in each case.

replicate	*expression level (arbitrary units)*				*normalised expression level: ratio/M*		
	M	**H**	**A**	**Q**	**H/M**	**A/M**	**Q/M**
a	1.30	2.19	24.45	14.15	1.68	18.77	10.86
b	0.46	2.25	24.39	19.64	4.93	53.60	43.17
c	9.29	6.53	64.23	38.70	0.70	6.91	4.17
d	3.87	3.60	35.65	30.89	0.93	9.20	7.97
e	10.19	6.65	55.60	60.19	0.65	5.46	5.91
f	4.36	10.94	172.34	21.06	2.51	39.52	4.83
g	17.12	30.76	388.48	173.45	1.80	22.70	10.13
mean	6.66	8.99	109.31	51.16	1.89	22.31	12.43
variance	34.83	101.66	1.78×10^4	3.15×10^3	2.26	330.16	190.19
s.e.	2.23	3.81	50.39	21.20	0.57	6.87	5.21

Table 14

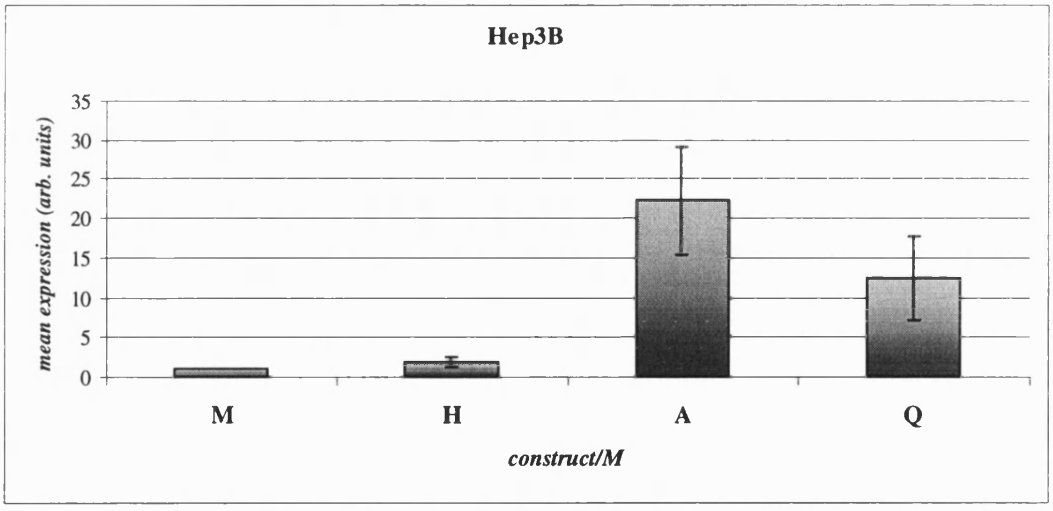


Figure 20.

Table 15. Summary of transient transfection data for undifferentiated *C2* cells. The ***M*** construct was treated as providing the basal level of expression in this experiment, to which all other constructs were compared. A normalised value of test construct expression was calculated as above (***Table 14***). The mean, variance and standard error of the mean (s.e.) were calculated for each construct, and each normalised construct, where $n=8$. There was no significant upregulation of gene expression from any construct over basal levels.

Figure 21. Expression levels of each test construct (***H, A*** and ***Q***) following transient transfection into undifferentiated *C2* cells. All test constructs were normalised by dividing by the level of basal expression, which is judged as expression levels following transfection of the ***M*** construct (see ***Table 15***). Error bars show the standard error of the mean in each case.

replicate	*expression level (arbitrary units)*				*normalised expression level: ratio/M*		
	M	**H**	**A**	**Q**	**H/M**	**A/M**	**Q/M**
a	2.84	0.19	4.49	3.88	0.07	1.58	1.36
b	16.82	41.72	42.65	44.31	2.48	2.54	2.63
c	76.56	72.13	67.55	92.86	0.94	0.88	1.21
d	28.42	29.98	27.55	16.27	1.05	0.97	0.57
e	0.37	0.65	1.09	1.06	1.75	2.93	2.86
f	0.79	0.81	2.00	0.70	0.96	2.54	0.91
g	26.06	49.70	40.17	22.25	1.91	1.54	0.85
h	13.03	26.32	25.38	2.47	1.98	1.91	0.19
mean	20.64	27.69	26.36	22.98	1.40	1.86	1.32
variance	629.59	696.30	552.21	1020.95	0.59	0.57	0.91
s.e.	8.87	9.33	8.31	11.30	0.27	0.27	0.34

Table 15.

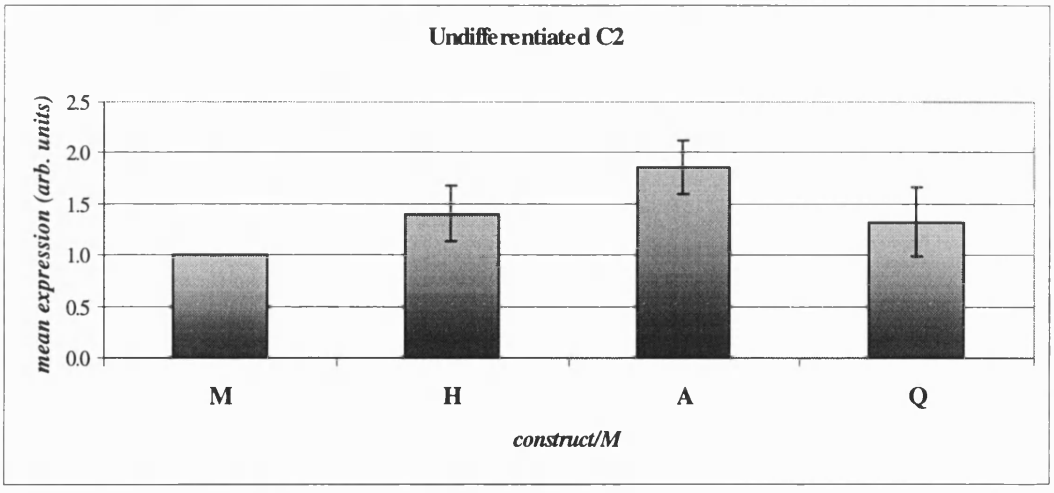


Figure 21

Table 16. Summary of transient transfection data for differentiated *C2* cells. The ***M*** construct is treated as providing the basal level of expression in this experiment, to which all other constructs are compared. A normalised value of test construct expression was calculated as above (***Table 14***). The mean, variance and standard error of the mean (s.e.) were calculated for each construct, and each normalised construct, where *n=4,* except in the case of a single ***H*** sample (sample *d*, nd) which was not recovered.

Figure 22. Expression levels of each test construct (***H, A*** and ***Q***) following transient transfection into differentiated *C2* cells. All test constructs were normalised by dividing by the level of basal expression, which is judged as expression levels following transfection of the ***M*** construct (see ***Table 14***). Error bars show the standard error of the mean in each case.

replicate	*expression level (arbitrary units)*				*normalised expression level: ratio/M*		
	M	**H**	**A**	**Q**	**H/M**	**A/M**	**Q/M**
a	112.23	79.95	60.86	87.28	0.71	0.54	0.78
b	22.79	14.08	16.05	9.25	0.62	0.70	0.41
c	5.58	5.33	2.98	4.57	0.96	0.54	0.82
d	1.61	nd	0.60	0.49	nd	0.37	0.30
mean	35.54	33.12	20.12	25.40	0.76	0.54	0.58
variance	2.70×10^3	1.66×10^3	783.59	1.72×10^3	0.03	0.02	0.07
s.e.					0.10	0.07	0.13

Table 16

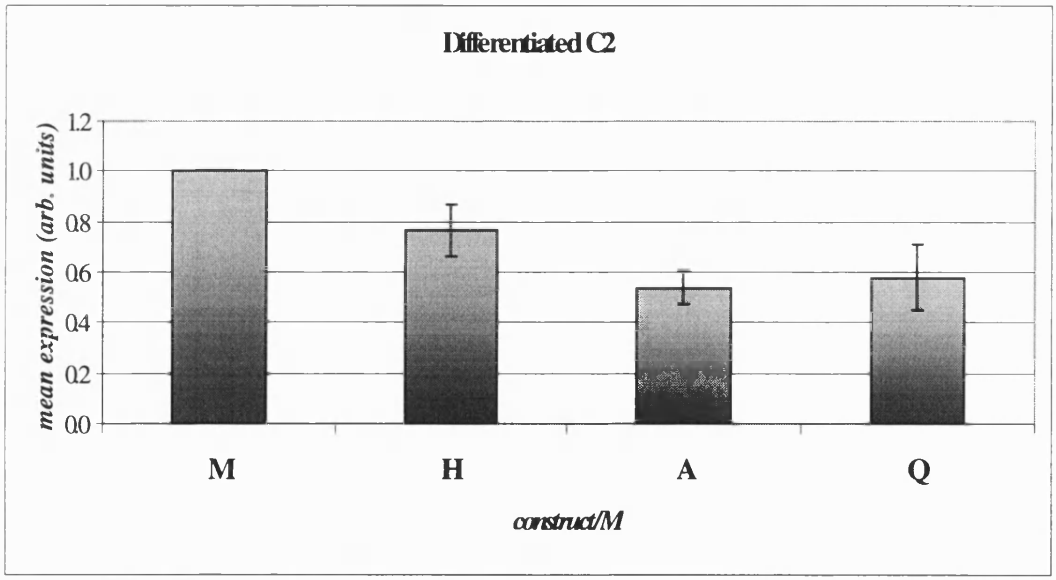


Figure 22

Conclusions

P3-*luciferase* reporter constructs were transiently transfected into four cell lines derived from different tissues; *Hep3B*, the human hepatocarcinoma cell line, *C2* a murine myoblast cell line, *Cos7* a monkey kidney cell line and *NIH 3T3*, a mouse fibroblast cell line.

In *Hep3B* cells, the ***H*** construct did not display any upregulation of gene activity over the ***M*** construct, which suggests that the CCD is unable to act as an enhancer in liver. This result is unsurprising, since the *H19* enhancers are thought to be the major enhancers for the *Igf2* and *H19* genes in endodermally-derived tissues [104]. Consistent with this is the enhancement of *luciferase* expression over basal levels observed with the ***A*** construct. In this case the expression of *luciferase* was increased approximately 20-fold over the basal expression from P3 alone when the *H19* enhancers were present. This figure compares favourably to the study of the activity of the *H19* enhancers in *Hep3B* cells performed previously [103], which reported an approximately 70-fold enhancement of gene expression from the *H19 promoter* when these enhancers were *in cis*. The difference between a ~20-fold enhancement compared to a ~70-fold enhancement may be a result of a number of factors, including a difference in RNA stability between the *H19* reporter and *luciferase,* in the relative affinity of the two promoters for interaction with the enhancer elements; or could merely be a result of different experimental conditions.

When the CCD and the *H19* enhancers were present in the same reporter construct (***Q***), reporter gene levels were reduced approximately 2-fold with respect to the *H19* enhancers alone. The downregulation of the *H19* enhancers in the presence of the CCD could be the consequence of a less optimal spacing between the enhancers and the promoter, rather than due to CCD-specific sequences. This possibility can only be ruled out by examining expression levels from a control construct comprising of P3-*luciferase*, the *H19* enhancers and an additional 2kb of non-coding DNA in place of the CCD. *In vivo* (**Chapter 4**) the CCD appeared to upregulate reporter gene expression in embryonic body tissues when *in cis* to the *H19* enhancers. This contradiction cannot be resolved without further experiments examining the effect of spacing on *H19* enhancer function, both *in vivo* and *in vitro*.

The CCD provided a candidate for the mesodermal-specific enhancers of *Igf2* and/or *H19*. With this in mind, reporter constructs were transiently transfected into the myoblast cell line *C2*. No upregulation of gene expression over basal levels was observed

with any of the reporter constructs, even when the cells were induced to undergo differentiation, a process known to lead to abundant expression of *Igf2* in this cell line.

This result correlates very well to the parallel studies of *in vivo* expression of P3-*luciferase* reporter constructs presented in **Chapter 4**. ***A, H*** and ***Q*** transgene-bearing mice expressed *luciferase* at low levels in skeletal muscle at D1, with no apparent differences in expression levels between constructs.

Preliminary experiments were conducted to establish whether the reporter constructs were expressed in further cell lines. In both *Cos7* cells and in *NIH 3T3* cells, no upregulation of reporter gene activity over basal levels was observed following transient transfection of any of the test constructs. Despite small sample sizes, it was concluded that neither the *H19* enhancers, nor the CCD could drive reporter gene expression in these cell lines.

In conclusion, P3-*luciferase* constructs containing the *H19* enhancers, in *Hep3B* cells, behave in a manner expected from previous transient transfection experiments [103], *in vivo* studies [104], and as discussed in this report (**Chapter 4**). The apparent down-regulation of these enhancers in the presence of the CCD cannot be verified without further experiments to confirm that CCD-specific sequences, rather than a generalised spacing effect is responsible.

In a cell line derived from embryonic muscle cell precursors, the *H19* enhancers do not drive *luciferase* expression from the P3 promoter, suggesting that these enhancers are not responsible for the abundant expression of *Igf2* and *H19* in skeletal muscle. This conclusion is supported both by *in vivo* studies of enhancer function in this work, and others [104].

The CCD does not drive reporter gene activity in any cell line examined, suggesting that it does not act as an enhancer of *Igf2* or *H19* in the mesoderm (*C2, NIH 3T3*), in liver (*Hep3B*) or in kidney (*Cos7*). This result is supported by the *in vivo* study of reporter gene activity presented in **Chapter 4**.

In order to dissect the function of the CCD further by a transient transfection method, a cell line expressing reporter constructs with the CCD *in cis* must be found. The major site of *luciferase* expression from CCD-containing transgenes was in the exchange tissues of the brain, so CCD function must be examined in a cell line derived from these tissues. The study of CCD function in a cultured cells derived from the choroid plexus is discussed in the next chapter.

CHAPTER 7, IS THE CCD ABLE TO DIRECT REPORTER GENE EXPRESSION IN AN *IN-VITRO* SYSTEM? PART 2, PRIMARY CULTURE.

Introduction

Experiments utilising a transgenic approach (see **Chapter 4**) have demonstrated that gene expression in the choroid plexus can be conferred on a reporter gene when *in cis* to the CCD (previously defined by a ~2kb DNA fragment, in **Chapter 1**). In order to further delineate activity in this region, in terms of discovering the specific sequences necessary for choroid plexus expression and the *trans*- acting factors that bind to these sequences within the CCD, an *in vitro* approach has been attempted. *In vitro* systems have previously been used successfully to isolate both enhancer sequences [103], and imprinting control regions [169], [170], [155] at the *Igf2/H19* locus. These studies made use of previously characterised stable cell lines in which to measure changes in reporter gene activity in the presence of candidate control elements. The CCD does not appear to drive reporter gene expression in a variety of stable cell lines (see **Chapter 6**), suggesting that it does not act promiscuously as a control element, and instead its activity could be limited to a small subset of tissues. The transgenic mouse studies presented in **Chapter 4** strongly suggest that the CCD can drive reporter gene expression in the exchange tissues of the brain, within which the choroid plexus is included.

A single stable cell line exists which originates from choroid plexus tissue, the *SCP* cell line [215]. This cell line was not used in this study for a number of reasons. Firstly, the cell line is infected with the Maedi-Visna virus [216], providing problems for containment, and raising doubt upon possible extrapolations from this highly unphysiological culture system to any function *in vivo*. Secondly, the *SCP* cells are fibroblastic in morphology, not epithelial, as would be expected from cells of the choroid plexus, suggesting that they may be derived from the mesenchyme underlying the choroid plexus, or that the cells may have transformed in culture. Thirdly, the CCD region under study is derived from the mouse, whereas *SCP* cells are ovine in origin. While the CCD has been shown to be conserved in a number of mammalian species [151], this conservation has not been demonstrated in sheep. The poor resemblance of *SCP* cells to the physiological structure from which they were

derived was judged to render *SCP* cells an inadequate model in which to study CCD function *in vitro.*

Instead, a strategy has been adopted to derive primary cultures of choroid plexus epithelium in which to test reporter gene expression in the presence of the CCD. Primary cultures of choroid plexus epithelium have been derived successfully from a number of mammalian species [192, 217, 218, 219, 190, 220], including the mouse [221]. A novel method of primary culture of choroid plexus cells in the mouse, and characterisation of these cells is described below.

Results

Culture of CP cells

Dissection and cell dispersal

The method of mouse choroid plexus primary culture described by Thomas et. al. [221] differed from those protocols described for other mammalian species in that embryonic (e12.5) choroid plexuses were isolated for culture rather than post-partum samples. Due to this, more than 40 embryos were required to achieve a reasonable density of cells for culture. As primary cultures from 4-6 week old rat choroid plexuses (with 10 animals required per culture) are successful [220], it was decided to attempt to culture post-partum mouse choroid plexus. This would allow both the minimisation of the number of animals required for these experiments, and reduction of the length of the dissection protocol. The age of 7-10 days post-partum (D7-10) was used, as mouse choroid plexus is known to stop proliferating by 2-weeks after birth [222], and the success of primary culture from rat cells decreased when animals were older than weaning age [220] (approximately 21 days in the mouse).

The choroid plexus was dissected from the lateral and fourth ventricles of approximately ten (usually one litter) D7-10 mice. The dissected tissue was treated with the protease trypsin to release the cells from surrounding connective tissue (as described in several previous works [217, 190, 219]). Approximately $1x10^6$ cells were liberated per experiment, but the number of viable cells was very low. Crook et. al. [192] tested several proteolytic enzymes for their ability to release viable cells from bovine choroid plexus tissue. A comparison of collagenase, dispase, pronase and the combination of collagenase

and dispase revealed that treatment with pronase generated the highest number of viable cells, and also the most effective cell attachment. The dissection was repeated using pronase to disperse the choroid plexus cells in favour of trypsin. Pronase digestion of mouse tissue effectively released the cells, and gave a good level of cell viability (an average of 5.4 $x10^6$ viable cells per experiment/ ~80% of the cells dispersed, n=13). Pronase treatment must be well controlled, since prolonged digestion significantly reduced cell viability (not quantitated).

Culture Conditions

As previously reported, choroid plexus epithelial cells do not attach, or grow well on a plastic substrate [221]. Attempts to grow the dispersed cells on untreated culture vessels confirmed this report. Previous investigators have successfully cultured choroid plexus epithelium on a variety of substrates (Matrigel and collagen I [221], laminin [219, 218]). *In vivo*, the choroid plexus epithelium overlies, and is supported by a basement membrane provided by fibroblastic mesenchymal tissue [223]. A basement membrane provided by fibroblasts grown in culture might then be a good approximation of the extracellular matrix (ECM) that the choroid plexus epithelium contacts *in vivo*. This ECM was obtained by the method of Freshney [191]. The murine fibroblast cell line *NIH3T3* was grown to a confluent monolayer, and then the cells lysed by the addition of a detergent. The contents of the cells was then washed away, leaving a residue of the ECM laid down by the fibroblasts on the base of the culture dish. The dispersed cells were plated onto culture dishes treated in this manner, and the majority of cells attached (not quantitated) within 48hrs. The cells were then washed to remove unattached cells and debris.

Several days after plating the cells were predominantly of two morphological types: flat, polygonal closely opposed cells that formed islands; and fibroblastic cells that were dispersed among the islands. The morphology of the polygonal cells was similar to that of several epithelial cell lines in culture (see below, and Freshney, (1987) [191]). Fibroblasts initially comprised only a small proportion of the cultured cell population, but they rapidly proliferated to occupy the space among the polygonal cell islands. The aggressive growth of the fibroblastic cells thus prevented further growth of the epithelial cells. In order to prevent the growth of the fibroblast population, the culture medium was supplemented with cytosine arabinoside. The addition of this reagent at an early stage in the culture (48hrs) was sufficient to prevent the fibroblasts from proliferating, and has been reported

to have no influence on the growth of epithelial cells [190]. With this treatment, the epithelial cells continued to grow and divide until confluency was reached, usually within 10 days.

Characterisation of primary culture cells

In order to verify that the cells grown in culture are in fact derived from the choroid plexus, two criteria were specified: Firstly, do the cells isolated in primary culture display the morphological characteristics of epithelial cells? Secondly, do the cells express choroid plexus-specific marker genes?

Morphology

The cells grown in primary culture display several characteristics of epithelium as can be visualised by light microscopy. As shown in ***Figure 23a*** the cells display a 'cobblestone' morphology of polygonal cells that is a characteristic of epithelial cells grown in a monolayer. At higher magnification (***Figure 23b***), the cells appear to be have a rough texture, which is indicative of the presence of villi on the apical surface of the cells, another characteristic of epithelium.

Gene expression

Two genes were chosen to act as markers for choroid plexus epithelium in this study, the *transthyretin* gene (*TTR*) and *Igf2*. As has been discussed at length in the **Introduction**, the *Igf2* gene product is absent from all brain tissues, excepting the choroid plexus and leptomeninges [97], and may therefore serve as a marker for these tissues. In addition, as the purpose of this study is in the elucidation of regulatory factors for *Igf2* in the choroid plexus, it is of obvious interest to establish that this gene is expressed in the culture system.

TTR, or *prealbumin* plays an important role in the plasma transport of vitamin A and is also involved in the transport of thyroid hormone. Its major sites of expression in rodents and man is in the liver, visceral yolk sac and the choroid plexus [224, 225]. In the developing rodent brain, *TTR* is expressed from the earliest stages of choroid plexus development. Transcripts can be detected in the choroid plexus primordium, a layer of epithelial cells associated with two distinct areas of thickened mesenchyme which overlie the neural tube at the midline [226]. *TTR* is thus an early marker for choroid plexus epithelium.

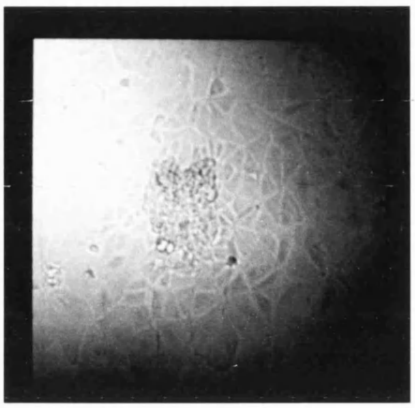


Figure 23a. Low power view (x100) of choroid plexus primary cultures. The cells display a cobblestone morphology characteristic of epithelial cells.

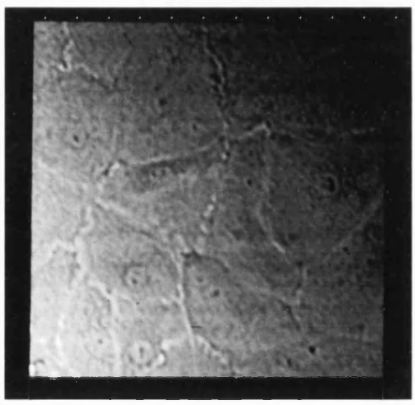


Figure 23b. High power view (x400) of choroid plexus primary cultures. The granulated appearance of the surface of the cells is due to the presence of villi.

Primary culture cells were fixed and stained with an anti-*TTR* primary antibody, and the staining pattern visualised by the use of a fluorescein-labelled secondary antibody. The resulting staining pattern is shown in ***Figure 24***. The *TTR* protein is present at high levels in the cultured cells, and shows a cytoplasmic distribution.

To verify that both *TTR* and *Igf2* genes are expressed in the primary cultures, RNA was isolated from these cells, and from cells of the human hepatocyte cell line *HepG2*. *HepG2* cells would also be expected to express both markers, so this sample serves as a positive control. The RNA was subjected to Northern blotting, and hybridised with probes to exons 1 and 2 of the *TTR* gene, and exon 6 of *Igf2* (see **Materials and Methods**). ***Figure 25*** shows that both the 3kb *Igf2* transcript, and *TTR* transcripts are present in the RNA samples derived from choroid plexus primary cultures. *TTR* mRNA can be visualised as two transcripts, one at 0.7kb, which is the expected transcript size as reported by previous work [227]. The second fragment at approximately 1kb is compatible with the report of a choroid plexus-specific *TTR* transcript reported in mouse by Kita et. al. [228]. *HepG2* cells express only the 0.7kb *TTR* transcript. These cells did not show hybridisation to the *Igf2* probe. This may be due to the failure of the *Igf2* probe derived the mouse to label to the human transcript under the stringent hybridisation conditions.

The primary cultures derived from mouse choroid plexi thus appear to have retained their identity, both in terms of their morphology, and by the expression of the choroid plexus-specific marker genes *TTR* and *Igf2*.

Transient transfection of choroid plexus primary cultures

CCD deletion constructs

Three of the constructs that were transiently transfected into choroid plexus primary cultures (***M***, ***H*** and ***A***, described in ***Figure 3***, **Chapter 1**) have previously been assayed for expression in cultured cells, and the results of this analysis are discussed in **Chapter 6**. Two additional expression constructs (***pCCD11a*** and ***pCCD4a***) were created and transiently transfected. These constructs were derivatives of the ***H*** construct, and are depicted in ***Figure 26.*** The constructs were created in order to assess the role of the DNaseI hypersensitive sites at the CCD in the activity of this region. As will be discussed in **Chapter 8**, the DNaseI hypersensitive sites, which initially characterised the CCD

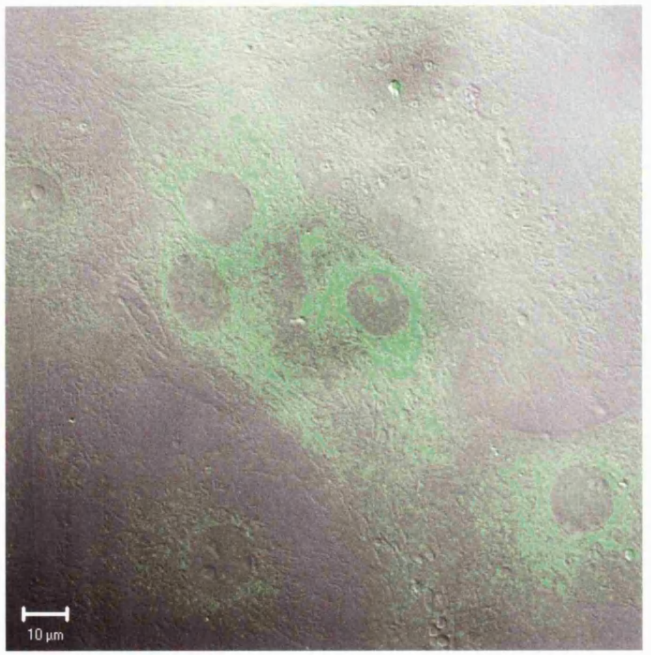


Figure 24a.
Fixed primary culture cells treated with an anti-TTR preparation followed by a fluorescein-labelled secondary antibody. Cells are viewed under FITC, and the image overlaid with a bright field image of the cells. Cytoplasmic staining for TTR can be clearly seen.

Figure 24b.
Fixed primary culture cells treated with the secondary antibody only, and viewed under the same conditions as the cells in Figure 24a. There is no background staining in the absence of the primary antibody.

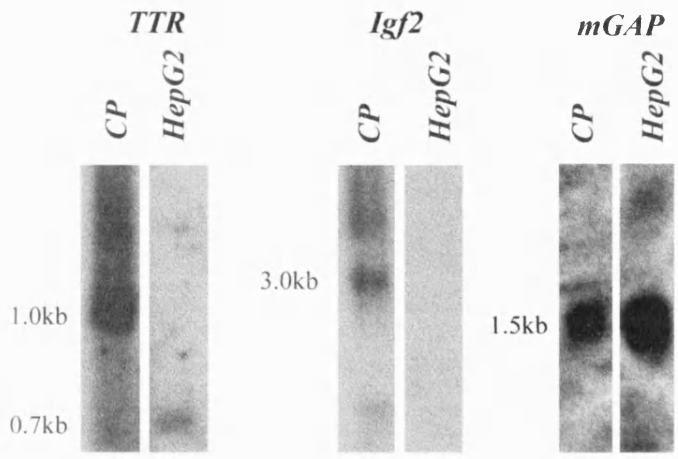


Figure 25.
Northern blot analysis of transthyretin and Igf2 gene expression from cultured cells. RNA was extracted from choroid plexus primary cultures (CP) and the hepatocyte cell line HepG2, and analysed with probes for TTR and Igf2. The TTR probe hybridises to a transcript of 0.7kb in both CP and HepG2 samples, in addition to a 1kb transcript in the CP sample. A 3kb fragment that corresponds to the Igf2 transcript is visible in the CP, but not the HepG2 sample. Hybridisation to the mGAP probe reveals that abundant RNA is present in both samples.

[151], have been mapped to the 5' end of this region [229]. It was of interest to compare the activity of the CCD when the hypersensitive sites were present, to when they were deleted. ***pCCD11a*** (***Figure 26b***) is comprised of the 1.1kb region containing the hypersensitive sites, fused to the most 3' ~100bp of the CCD. A 0.8kb fragment is deleted from the middle of this region. ***pCCD4a*** (***Figure 26c***) contains the 3' 0.9kb of the CCD. A 1.1kb region containing the hypersensitive sites has been deleted. These two deletion constructs are of a similar size (0.9kb and 1.2kb, respectively), so any differences between them are likely to be a result in differences in DNA sequence, rather than the length of the transfected plasmid.

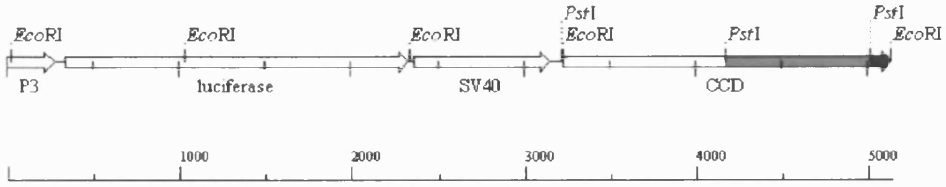
Transfections

Construct DNA was transiently transfected into a choroid plexus primary culture cells, to allow examination of *luciferase* activity when P3 is *in cis* to the *H19* enhancers (***A*** construct), the CCD (the ***H*** construct), and the truncated CCD in the presence (***pCCD11a***) and absence (***pCCD4a***) of a 1.1kb region containing DNaseI hypersensitive sites. Levels of expression from P3 alone (***M*** construct) was measured in all experiments, and provided a baseline to which other construct activity was compared. In all cases a construct containing the *LacZ* gene driven from the *SV40* promoter was co-transfected with test constructs to control for transfection efficiency, and levels of total protein were measured to control for cell number. All figures of reporter construct activity are calculated as ng *luciferase*/mg total protein/level of *β-galactosidase* activity. When no *luciferase* constructs were present, expression was never detected over background levels for the *luciferase* assay.

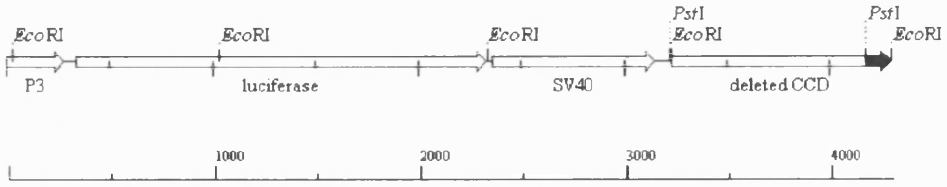
Luciferase expression levels following transient transfection of construct DNA into choroid plexus primary cultures are tabulated below, and presented as ratios of test construct to the basal expression level (***Table 17***). The ratios of mean test construct expression over basal expression are also presented in graphical form (***Figure 27***). The transfection experiments were performed at least four times for each expression construct. It should be noted that in all cases the transfection efficiency (as assayed by the level of *β-galactosidase* activity) was very low, and that this may contribute to the variation in expression levels observed between experiments, as well as increasing the possibility of artefactual results.

The expression levels in choroid plexus cells do not differ in magnitude between the test constructs ***H*** and ***A***, and basal levels (***M***). In the cells expressing the reporter constructs, the addition of the CCD or the *H19* enhancers has no effect on reporter gene

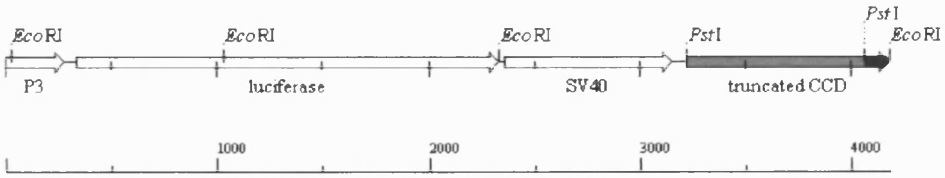
Figure 26. Cloning of CCD deletion constructs



a) Restriction map of the H construct showing the CCD divided into three regions delineated by EcoRI/PstI restriction sites.



b) pCCD11a: a CCD deletion construct containing the most 5' 1.1kb of the CCD (which includes the DNaseI hypersensitive sites) fused to the ~100bp most 3'. A ~0.8kb region is deleted from the middle.



c) pCCD4a: a CCD deletion construct containing the 3' 0.9kb of the CCD. The 1.1kb 5' region containing the DNaseI hypersensitive sites has been deleted.

expression. In the case of the CCD deletion constructs, ***pCCD11a*** does not appear to differ significantly from the parent construct ***H*** in the level of reporter activity from P3. However, when the DNaseI hypersensitive sites are removed from the CCD region, the element appears to act as a silencer, depressing reporter gene expression to a level below that of the basal ***M*** construct. The small number of replicates make it impossible to assign a statistical significance to this result, however.

Creation of a cell line

The primary cultures of choroid plexus cell produced in this study generally achieved confluency as a monolayer and were unable to proliferate further. In order to establish a cell line from these primary cultures it was necessary for the cells to survive passage once they had reached confluency, in order for the culture to be split and repropagated. To this end, the cells were dissociated from their attachment to the culture flask by treatment with trypsin, and replated into new culture dishes (prepared as described previously with an ECM from dissaggregated *NIH 3T3* cells). The number of viable cells was counted following trypsinisation, and the ratio of cell survival calculated by division of this number by the number of cells originally plated. Following trypsinisation, the number of surviving cells decreased considerably, and the majority of cells did not reattach to the new culture vessel. Those cells that did reattach within a few days displayed the 'rounded up' morphology of apoptotic cells, and did not subsequently proliferate. Much of the loss of cell viability could be accounted for by the harsh treatment of trypsinisation. However, the fact that viable cells were recovered after this treatment, but then failed to proliferate further, suggests that these cells had reached the limit of their proliferative potential. Cells derived from the choroid plexus of normal mice did not retain their proliferative potential beyond passage.

In an attempt to increase the proliferative potential of the choroid plexus cells, with the aim of establishing immortal cell cultures, the experiments were repeated with the choroid plexi of mice which lacked the tumour suppressor gene *p53*. Two lines of evidence suggested that cells lacking the *p53* gene might be competent to produce an immortal choroid plexus cell line. Firstly, primary embryonic fibroblasts (PMEFs) derived from *p53* null mice are immortal [230], and can be maintained indefinitely in culture. Secondly, a strain of transgenic mice have been generated that express the immortalising *SV40 T–antigen* in choroid plexus epithelium (the T_{121} strain [231]). The variant of *SV40 T–antigen*

replicate	*normalised expression level: ratio/M*			
	H	***A***	*p*CCD***4a***	*p*CCD***11a***
a	0.230	0.520	*nd*	*nd*
b	1.070	3.130	*nd*	*nd*
c	0.810	1.040	*nd*	*nd*
d	0.699	0.397	0.130	1.129
e	*nd*	*nd*	0.096	1.807
f	1.585	0.807	0.108	0.922
g	0.467	0.604	0.129	0.387
mean	**0.81**	**1.08**	**0.12**	**1.06**
variance	**0.48**	**1.03**	**0.02**	**0.587**
s.e.	**0.28**	**0.41**	**0.06**	**0.383**

Table 17. Summary of transient transfection data for choroid plexus cells. Constructs were transiently transfected into the cultured cells. The ***M*** construct is treated as providing the basal level of expression in this experiment, to which all other constructs are compared. A normalised value of test construct expression is presented as the ratio of test construct expression vs. basal (***M***) expression. The mean, variance and standard error of the mean (s.e.) are calculated for each normalised construct (where *n=6* for ***H/M*** and ***A/M*** and *n=4* for ***pCCD4a/M*** and ***pCCD11a/M***). *nd* indicates that the construct was not tested in this experiment.

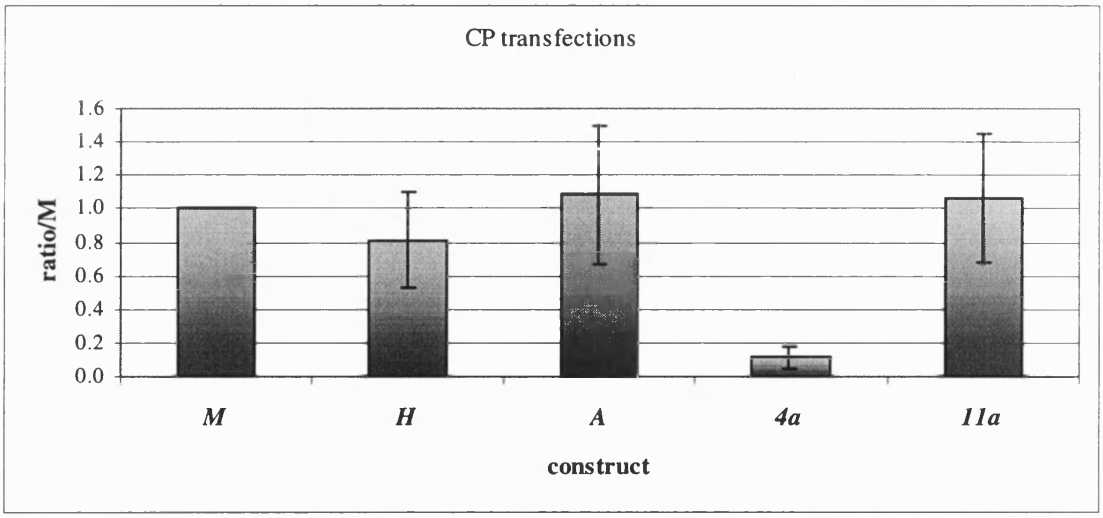


Figure 27. Expression levels of each test construct (***H, A***, ***pCCD4a*** (***4a***) and ***pCCD11a*** (***11a***)) following transient transfection into choroid plexus (CP) cells. All test constructs are normalised by dividing by the level of basal expression, which is judged as expression levels following transfection of the ***M*** construct (see ***Table 17***). Error bars show the standard error of the mean in each case.

expressed by T_{121} mice is truncated, so that it inactivates the *retinoblastoma* family proteins, but not *p53*. Choroid plexus epithelium usually withdraws from the cell cycle within two weeks after birth and remains in a quiescent state. In T_{121} mice the cells are induced to proliferate aberrantly, and the animals develop choroid plexus tumours within 30-40 weeks of life. However, tumour growth in T_{121} choroid plexus is suppressed by the induction of *p53*-dependent apoptosis. If T_{121} mice are crossed onto a *p53* null background, choroid plexus tumours develop within 4-5 weeks of life [222]. *p53* loss of function is thus a prerequisite for the immortalisation of choroid plexus cells.

Mice with a null allele of *p53* (described by Clarke et. al. [180]) were bred to homozygosity, and the offspring of homozygous null parents used in the creation of choroid plexus primary cultures. The resultant morphology was identical to that of primary cultures derived from wild type mice, and the rate of growth was not obviously different, though no quantitative comparison was made. On reaching confluency the cells were passaged, and the survival ratio calculated. One week later the cells were passaged for a second time, and the survival ratio calculated again. The results of these experiments, for three independent preparations of choroid plexus cells, are shown in ***Figure 28***. The small number of data points makes this work preliminary, but a clear trend can be seen. After the first passage the cells had in two out of three instances reduced in number to less than 50% of their plating number. By the second passage, the number of surviving cells was very low in all three experiments, and in all cases by a week following the second passage the cells had rounded up and ceased proliferating. While a direct comparison between wild type and *p53* null cells was not attempted, it appears that *p53* null cells have some survival advantage over their wild type counterparts. Despite this, the lack of *p53* alone was not sufficient for immortalisation of primary choroid plexus cultures.

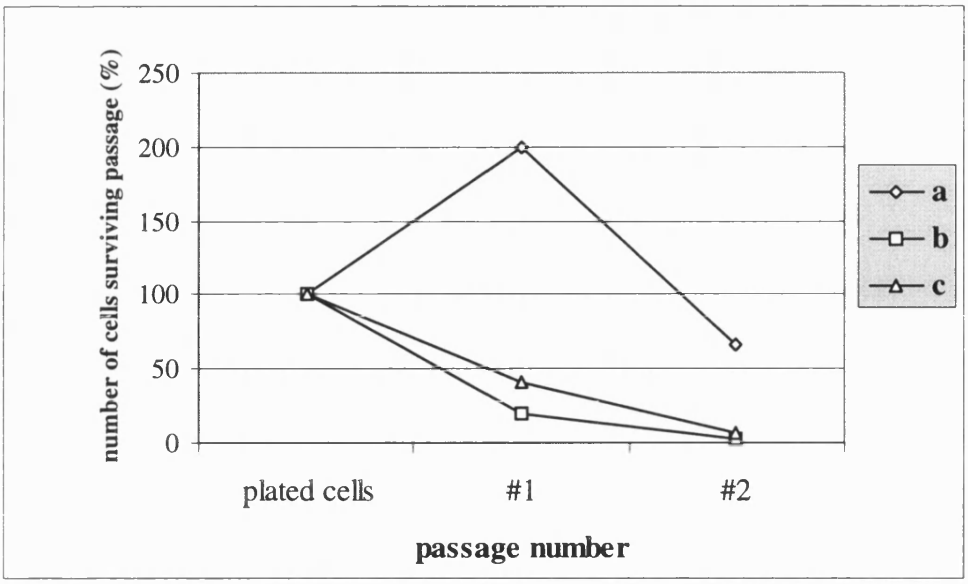


Figure 28. Survival of *p53* null choroid plexus cells following passage. Three independent preparations of choroid plexus cells (**a, b** and **c**), were counted prior to plating into the culture flask (plated cells), disaggregated when the cells reached confluency (#1), and the cells subsequently replated. The third count (#2) was taken 7 days following this passage, and the cells again replated. Within a week of the third count, all three preparations consisted exclusively of rounded up, non-proliferative cells.

Conclusions

When placed *in-cis* to *Igf2* promoter 3, the CCD can drive reporter gene expression in the brains of transgenic mice (Ward et. al. [108], and **Chapter 4**). This activity is limited to the exchange tissues of the brain, i.e., the choroid plexus and leptomeninges, sites of endogenous *Igf2* gene expression [70, 97]. In order to identify the sequences within the CCD that are responsible for this enhancer activity, an *in-vitro* approach has been attempted. The CCD is unable to upregulate reporter gene activity in a range of immortal cell lines that are representative of tissues other than choroid plexus (see **Chapter 6**). This element does not therefore act to confer a tissue-independent upregulation of P3. As the CCD drives reporter gene activity in the exchange tissues of the brain, an *in-vitro* system was sought that is representative of this tissue. No such cell line was available at the time this study was undertaken, so instead a strategy was adopted to derive primary cultures from choroid plexus tissue.

Choroid plexus primary cultures have been successfully derived from a number of mammalian species including pig [190], rat [217, 218, 220], rabbit [219], cow [192] and mouse [221]. In this study a novel method of primary culture of murine choroid plexus cells was developed. Previously reported primary culture methods for mouse cells [221] utilised embryonic tissue, which necessitated the use of a large number of animals per experiment (>40). The cell culture protocol developed here is the modification of a method used for the culture of porcine cells [190]. Three important modifications to this basic protocol allowed successful primary culture of mouse choroid plexus tissue. Firstly, the dissected tissue was dissaggregated with the protease pronase, which improved both cell viability, and increased the number of cells that attached successfully to the culture dish. Secondly, the cells were grown on a basement substratum that had previously been laid down by a confluent layer of fibroblasts. This modification greatly increased the plating efficiency of the dispersed cells, and has not previously been reported in a choroid plexus primary culture protocol. Thirdly, the nucleotide analogue, cytosine arabinoside, was added to the culture medium to prevent the growth of contaminating fibroblasts. These three modifications provide a robust protocol for the isolation and culture of choroid plexus epithelium.

The identity of the cultured cells was verified by examination of cell morphology, and by the study of gene expression. The cultured cells display a flattened polygonal

morphology characteristic of epithelium. At high magnification the surface of the cells appears rough, which can be indicative of the presence of villi. However, this assumption could only be verified by examination of the cells by electron microscopy. The expression of two genes, *transthyretin* and *Igf2* was detected in the cultured cells. *Transthyretin* is a highly specific marker for choroid plexus epithelium, as this is its only site of expression in the mammalian brain [226]. *Igf2* expression is also localised in the mammalian brain to the choroid plexus and leptomeninges [97]. The detection of transcripts from both genes in the cultured cells confirms their identity as choroid plexus epithelium.

Once the identity of the cultured cells as choroid plexus epithelium had been confirmed, the cultures were then utilised for the study of CCD function. The experimental method employed was transient transfection of *luciferase* reporter constructs into the cells, and subsequent assays for reporter gene activity in the presence of the test constructs. Unfortunately, the cells were largely refractory to transient transfection by the standard method used in this study (see **Materials and Methods**), as judged by the level of enzyme activity following the introduction of a *LacZ*-bearing reporter construct. As the resultant transfection efficiency of the test constructs was low, the data obtained for the expression of *luciferase* reporter constructs should be treated with caution, for they may be representative of only a small number of transfected cells in each experiment. The preliminary data did suggest, however, that neither the CCD nor the *H19* enhancers could drive gene expression above the level of expression obtained from the P3 promoter alone. This is a surprising result, as the assays on expression from these constructs *in-vivo* (see **Chapter 4**) indicated that both elements were capable of enhancing gene expression from P3 in this tissue. To test the possibility that multiple activities were present within the CCD region, which may complicate the analysis, *luciferase* reporter constructs were constructed that were internally deleted for the DNaseI hypersensitive sites, or for a deletion of 0.9kb at the *H19*-proximal end. When the 0.9kb fragment was deleted, reporter gene expression levels did not differ from levels of expression from P3 alone. However, deletion of the hypersensitive sites brought about gene expression at a level approximately 10-fold lower than from the promoter alone. As this experiment was only repeated 4 times, too much significance should not be attached to this result, but it may indicate that more than one activity is present within the CCD region, one which includes the hypersensitive sites, and a second region that can act as a silencer. If this result should prove repeatable, it would

indicate that when the region containing the hypersensitive sites is removed, the silencer is able to downregulate reporter gene activity.

In order to effectively test CCD action in choroid plexus cultured cells, it will be necessary to optimise the transient transfection protocol, possibly by utilising reagents designed for transient transfection of primary cultures, or by the creation of stable cell lines. Optimisation of this protocol will require a large number of cells, and therefore would be more sensibly conducted in a cell line, rather than in primary cultures. Trial experiments were conducted to obtain immortalised cells in culture, by creating cultures from choroid plexus tissue derived from *p53* null mice. Fibroblasts obtained from such animals are immortal [230], and loss of *p53* is known to play a role in tumourigenesis of the choroid plexus in a mouse model [222]. However, loss of *p53* alone was not sufficient to immortalise choroid plexus cells in this culture system. Inactivation of *p53* and the *retinoblastoma* family proteins by the expression of SV40 large T-antigen in the choroid plexus of transgenic mice leads these animals to develop choroid plexus tumours within 5 weeks of life [231]. Choroid plexus tissue derived from these mice would provide a good source material for the formation of an immortal choroid plexus cell line. Future work in the laboratory will seek to establish a mouse choroid plexus epithelial cell line from T-$_{121}$/*p53* null choroid plexus, utilising the primary culture protocol developed in this study. The development of such a cell line will be valuable not only in the study of the mode of action of the CCD, but also in the study of *Igf2* gene regulation in the choroid plexus. Choroid plexus cells could be grown *in-vitro* which are derived from mice that are a product of an F1 cross between mouse strains polymorphic for a marker in the *Igf2* gene. The parental origin of the two *Igf2* alleles could thus be distinguished in this system. *Igf2* expression would be expected to be biallelic in cells derived from choroid plexus tissue, and in the system described above, this supposition can be tested. It is not known at present how the *Igf2* gene escapes imprinting in the choroid plexus, and the small size of this organ in the mouse makes investigations in this tissue problematic. An *in-vitro* culture of choroid plexus tissue (either primary cultures, or an immortal cell line) could be utilised to study the allele-specific expression of *Igf2* in the presence of e.g. histone acetylating agents or DNA demethylating agents, or in cells derived from different developmental stages.

In summary, primary cultures have been derived from murine choroid plexus tissue, and the cells resemble their source tissue both in morphology and in the expression of marker genes. Technical problems prevented a rigorous examination of CCD function in

these cells, but transiently transfected expression constructs containing the CCD do not appear to express the reporter gene at levels above that from P3 alone. An immortal cell line established from these cultures would be a valuable resource in examination of the mode of biallelic expression of *Igf2*. While this study indicates that loss of *p53* is not sufficient for immortalisation of these cells, the loss of the *retinoblastoma* family protein function, as well as *p53* function by the expression of SV40 T-antigen in the choroid plexus may well provide starting material with which to establish a cell line.

CHAPTER 8, SEQUENCE COMPARISONS BETWEEN THE CCD REGION IN A NUMBER OF DIFFERENT MAMMALIAN SPECIES

Introduction

Cross hybridisation analysis of genomic DNA from a range of mammalian species (including human, cow and rat) to a mouse CCD probe demonstrated that sequences exist in each of these organisms that share a high degree of evolutionary conservation with the mouse CCD region [151]. While demonstrating that such sequences exist, this analysis does not identify if homologous CCD sequences are present in the context of the *Igf2* and *H19* genes in other organisms. Furthermore, the analysis does not identify which sequences within the 2kb CCD probe used are conserved among these organisms.

As discussed in the introduction, comparisons between the mouse and human *H19* DMD region at the sequence level revealed several short conserved sequences, which were subsequently found to bind the boundary element protein *CTCF* ([156], [169], [170]). Cross-species analysis of the nucleotide sequence of potential regulatory regions can therefore identify good candidates for protein binding factors, and thus indicate the function of such regions.

The aim of the work discussed in this chapter was to answer the following questions: i) Does the *Igf2/H19* loci in species other than the mouse contain a region homologous to the CCD? ii) Are specific sequence motifs within the CCD region common across a range of species, and if so, what can the nature of these sequences tell us of CCD function?

Results

Sequencing of the 3' mouse CCD

The original CCD clone described by Koide et. al.[151]) was a 2kb *EcoRI* fragment. Approximately 1kb of this clone had been sequenced, and the DNaseI hypersensitive sites mapped. An initial step in the analysis of this region was to obtain the full sequence of the CCD clone. The portion of the CCD with no sequence data (approximately 900bp) was subcloned into a plasmid vector (see **Materials and Methods**), and sequenced from

flanking primers in the vector sequence in both orientations. The resulting sequence information was assembled using the GELASSEMBLE algorithm of GCG (the Wisconsin Package Version 8.0). The complete sequence of the CCD clone used in this study was thus obtained, and is shown in ***Appendix 2***.

Evolutionary conservation between mouse and human

Comparison of the complete CCD sequence with previously published sequences was carried out using the National Centre for Biotechnology Information's (NCBI) sequence similarity search tool, BLAST. A significant match was found to a PAC (P1 Artificial Chromosome) genomic clone of human chromosome 11p15.5 (accession number AC004556), approximately 30kb upstream of the *H19* promoter. This match demonstrated that homologous sequences to the CCD exist in the human, at a comparable location, i.e., between the *Igf2* and *H19* genes.

The alignment of the mouse and human CCD as performed by BLAST is shown in ***Figure 29***. Over the entire CCD region, approximately 55% sequence identity is observed, however the majority of this similarity resides in two domains, indicated here as Region 1 and Region 2.

Region 1 extends from 425-542bp of the mouse sequence, and shows approximately 83% sequence identity between mouse and human, with no gaps. The DNaseI hypersensitive sites (by which the CCD was first characterised, [151]) have previously been mapped to this region in the mouse (Dr J. F-X. Ainscough, pers. comm.), and their position is noted in ***Figure 29.***

Region 2 extends from 1629-1879bp of the mouse CCD, and shows 87% sequence identity between the two species. Two gaps are required in the human sequence to give an optimal alignment. Such extensive sequence identity appeared unusual for a regulatory region, where sequences are conserved for short blocks. A further possible role for this region is that it encodes a gene, or part of a gene. In both the mouse and the human Region 2 of the CCD contains an extended region encoding no stop codons. In addition, the most extended open reading frame (ORF) in the mouse could initiate at the ATG at position 1617. This start codon is not conserved in the human sequence however. If Region 2 encodes only part of a gene, the sequence need not initiate at an ATG codon, as it may be spliced onto an initiating exon. If this is the case, a splice acceptor consensus sequence would be predicted to lie at the CCD-internal end of the conserved sequence. The entire 2kb mouse sequence was submitted to the Splice Site Prediction by Neural Network

software package at the Berkeley Drosophila Genome Project. This software predicted the existence of several splice acceptor consensus sequences within the CCD region. Of interest was a 41bp sequence at position 1613-1653 that demonstrated a high level of conservation to optimal mammalian splice acceptor sites (82% sequence identity). Repeating the procedure with sequence from the human CCD region revealed that the corresponding 41 nucleotides (according to the alignment in ***Figure 29***) would also be predicted to function as a splice acceptor site, though this sequence was less optimal (55% identity with the optimal splice acceptor sequence). The position of this putative splice junction is illustrated in ***Appendix 2***.

Cloning the rat CCD

The discovery of a long open reading frame in Region 2 in both mouse and human raised the possibility that a gene may reside at the CCD. Koide et. al. [151] investigated this possibility by performing hybridisation analysis to mouse RNA from a variety of developmental stages with a CCD probe, and found no evidence of a transcript. It is possible however that such a transcript is present at different developmental stages than those analysed, or expressed at levels below the limits of detection of their analysis.

Analysis of the nucleotide sequence of a candidate coding region can provide indirect evidence that such a region encodes a protein. Due to the redundancy of the genetic code, the third position of a codon specifying a given amino acid can vary. The first two nucleotide positions of a codon would therefore be expected to be more highly conserved than the third position, i.e., the rate of non-synonymous substitution (K_a) of a nucleotide sequence specifying an amino acid would be expected to be lower than the rate of silent, or synonymous substitutions (K_s). How much, then, would sequences be expected to vary between two organisms, both at neutral sites, and at functional sequences? McVean and Hurst [68] have compiled a database of the rates of synonymous and non-synonymous substitutions between 422 orthologous genes in the rat and mouse. From this, the distribution of the rate of protein evolution between the two species has been calculated. If the CCD contains a gene, the calculated K_a/K_s value for the predicted protein would be expected to lie within this distribution. In order to carry out such an analysis it was necessary to obtain the nucleotide sequence of the CCD region in the rat. This was achieved by adopting a polymerase chain reaction (PCR) strategy using degenerate primers generated by comparison of mouse and human sequences.

Figure 29. Regions of homology between human and mouse CCD

The mouse CCD region was sequenced and compared to the published human sequence of this region. Two areas of striking sequence conservation were revealed:

CCDP3

```
mouse: 425  aaatttt[shaded]ggttagcataacaaccatttggcg  484
           ||||||||||||||| |||||||||||||||||||  |||||  |||  |||||||||
human:92275 aaatttttgttgtctcctgcgcccccgtgaggtcagtgggcagcatgacacccatttggcc 92216

mouse: 485  actgaggtggggttcctgggggggatg[shaded]c  542
           ||||||||||||  ||||  ||||||  |||||||||  ||  |  |||||||||  |  ||
human:92215 actgaggtggggctcctaggggggggtgcaggtgggaagcgaccccagccccctcccgc 92240
```

Region 1. Contains the DNaseI hypersensitive sites (shaded). It shows 83% homology with the human sequence over 118 nucleotides. There are no gaps.

CCDP1

```
mouse: 1629  ctcagatctgacggggccagagccaggacagaggaagagggctgtacagccagcccagct 1688
           ||||||||||  |||||||||  ||  ||  |||||||||  ||||  ||  |||  |||||||||
human: 90770 ctcagatctgccggggccagggctggggcagaggaag-gggc-gtgcagtgagcccagct 90713

mouse: 1689  gggacagcaaacaaacctctggggtctctcctttcccttgcctgcccaccacaggggtct 1748
           |||||||||||||  |||||||||||||||||||||||||||||||||  ||||||  ||
human: 90712 gggacagcaaacaggcctctggggtctctcctttcccttgcctgcccactgcaggggggct 90635

mouse: 1749  ggctggggccagagctgggggagacttcctcgatttatctctgcagccggtggccagccc 1808
           |||  ||||||||  ||  |||||||||||||||||||||||||||||||  |  ||  |||||
human: 90652 ggccggggccagggccgggggagacttcctcgatttatctctgcagccagcggacagcct 90593

mouse: 1809  taagcctggaatcccgggtcaagctaaggtttgcctgccctgctgtgagcagcttggagc 1868
             ||||||||||||  ||||  |  |||  ||||||||  ||  ||||||||||||||||||  |||
human: 90592 acagcctggaatctggggttaggctgaggtttgcgtgtcctgctgtgagcagcttgcagc 90533
```

CCDP2

```
mouse: 1869  tggtcagaagg 1879
             ||  |||||||
human: 90532 aggacagaagg 90522
```

Region 2. Shows 86% sequence homology over 251 base pairs. There are two single nucleotide gaps inserted into the human sequence at the beginning of this region. Both in the mouse and human this region is predicted to encode an open reading frame, though no database matches to the resulting protein sequence have been found. Primers (*CCDP1, CCDP2* and *CCDP3*) designed to amplify the CCD region in the rat are illustrated by arrows above their respective sequence.

Isolation of the rat CCD

Primers for PCR analysis were generated in regions of the CCD which were known to be the most highly conserved (based on the human/mouse comparison). In order to obtain the maximum length of sequence, primers were designed to amplify from Region 1 to the terminal end of Region 2. The Primer Designer (Scientific and Educational Software, Version 2) package was used to identify candidate primer sites in the mouse sequence, then these sequences were modified at nucleotides that were not conserved between mouse and human to include degenerate nucleotides. The resulting primer pair, *CCDP2* and *CCDP3* would be expected in the mouse to amplify a region of approximately 1.5kb, containing the majority of nucleotides in Region 1 and all of the nucleotides in Region 2. An additional primer, *CCDP1* was designed at the internal end of Region 2, to be used with *CCDP2* in order to amplify only the most conserved region (approximately 0.2kb in size), should amplification of the longer fragment prove problematic. The position of these primers with reference to the mouse CCD is illustrated by shading in ***Figure 29***.

Genomic DNA was prepared from the liver of a Sprague-Dawley rat as and from mouse tissue, and PCR was carried out using primer pairs *CCDP1* and *CCDP2,* as well as *CCDP3* and *CCDP2*. An initial experiment (shown in ***Figure 30a***) demonstrated that both primer pairs were able to give rise to amplification of fragments of the correct size from both mouse and rat genomic DNA. Under these conditions, the mouse sequences were amplified more readily than rat sequences, and many contaminating fragments were present in each sample. Optimisation of the PCR protocol led to a very successful amplification of a 1.5kb fragment from rat genomic DNA with primers *CCDP3* and *CCDP2* (shown in ***Figure 30b***). This fragment was subsequently purified and inserted into a cloning vector.

Verification of the rat CCD clone

Southern blotting of mouse and rat genomic DNA with a mouse CCD probe after digestion with a panel of restriction enzymes revealed a number of restriction fragment length polymorphisms between the mouse and rat CCD regions. ***Figure 31*** shows the results of this analysis for digestion of rat and mouse genomic DNA with the enzymes *PvuII* and *BamHI*. The expected fragment sizes from restriction digestion with these enzymes is shown in ***Table 18***. The mouse CCD contains two sites for the enzyme *PvuII*, giving rise to a 0.9kb fragment when hybridised to the mouse CCD probe, and no *BamHI* sites, giving rise to large fragments due to digestion outside the CCD region. The rat CCD region lacks

one of the two *PvuII* sites, giving rise to a 2.5kb fragment, and contains an internal *BamHI* site, resulting in hybridisation of a 3.5kb and a 6kb fragment to the probe. Restriction mapping of plasmids containing both the mouse and the rat CCD fragments confirms the presence of these restriction sites (see ***Figure 32*** and ***Figure 33***, and ***Table 18***). The amplified 1.5kb fragment therefore originated from rat genomic DNA, and was not amplified as a result of contamination with mouse DNA.

Sequencing the rat CCD

The 1440bp rat CCD sequence was assembled from 11 fragments of approximately 600bp each, obtained from sequencing primed using external primers in the cloning vector, and internal CCD primers (the *T7, SP6, CCDP1* and *CCDP4* primers, see **Materials and Methods**). The majority of the CCD sequence was thus obtained from comparison of at least three independent sequencing reactions. The 1440bp consensus sequence was created by the GELASSEMBLE program of the GCG software package, and verified manually. The full sequence of the rat CCD is shown in ***Appendix 2***.

Evolutionary conservation between mouse and rat

The overall level of sequence conservation between the mouse and rat CCD is 83%. As with the mouse-human comparison, the majority of this sequence similarity is confined to two domains, corresponding to the previously described Region 1 and Region 2. Region 1 shares 86% sequence identity between mouse and rat, over 93 nucleotides. The 93bp represents approximately $^{3}/_{4}$ of Region 1, the rest missing due to the design of the primers for PCR amplification. Region 2 shares 97% sequence identity between mouse and rat, indeed only 8 of a possible 251 nucleotides differ between the two species. As with the human and mouse CCD, the internal end of Region 2 (from nucleotides 1175-1219) contains a splice acceptor sequence (79% identity to the optimal mammalian consensus sequence) A summary of the comparative sequence data between the rat and mouse and the human and mouse CCD sequences is presented in ***Table 19,*** and a full alignment between the three species is shown in ***Appendix 2***.

Molecular evidence for a coding region at Region 2 of the CCD

The rat sequence, as well as the mouse and human, contains a long ORF at Region 2 of the CCD. The mouse and rat sequences also share a conserved ATG at the 5' end of this ORF.

The putative coding sequence was translated into an amino acid sequence (using the TRANSLATE algorithm of the GCG package), using the conserved ATG as a

convenient start position, and this protein sequence was used as the basis for the calculation of the rates of protein evolution discussed below. The following summarises an analysis of this putative protein coding region at the CCD conducted by L. Hurst (University of Bath), and is illustrated in ***Figure 34***. The putative coding region was found to have a rate of protein evolution (K_a) of 0.023 per site. This figure is at the low end of the distribution of the rates of protein evolution assembled in a mouse-rat comparison of 422 genes (as discussed above, where the mean was 0.036), and approximates the average for imprinted genes (K_a=0.0224, $\pm$0.00487, n=15). Therefore, the rate of evolution of this region is what would be expected of a protein coding sequence under stabilising selection. As the putative coding region, like many imprinted genes (including *Igf2*) [232], has a very low rate of evolution at silent sites (K_s=0.044 as compared to a mean of 0.178 for the 422 genes), the substitution rate of 5' flanking sequence was taken instead as the background rate of evolution in this region. With this modification, a K_a/K_s value of 0.114 is obtained, which is a little below the mean for the mouse-rat comparison (K_a/K_s=0.197).

The search for transcripts originating from Region 2

Despite good indirect evidence for an exon encoded by Region 2 of the CCD, no transcript corresponding to this sequence has been identified. mRNA from a range of foetal and neonatal mouse tissues have been analysed by Northern blotting, and ribonuclease protection analysis, using probes designed to be complementary to the putative coding region of the CCD (data not shown). While by no means exhaustive, these experiments did not uncover any reproducible evidence for a transcript originating from this region. Furthermore, searches of nucleotide and protein sequence databases for genes homologous to the putative gene have revealed no significant matches.

Figure 30a. Initial PCR amplification of the CCD region in mouse and rat (rat1 and rat4 denote independently prepared genomic DNA samples from rat tissue). Mouse sequences amplified more successfully than rat sequences. Many contaminating amplification products are present as well as the expected 1.5kb (with primers *CCDP2 +CCDP3*) and 0.2kb (with primers *CCDP1 +CCDP2*) products.

Figure 30b. Optimisation of the PCR protocol for amplification of the ~1.5kb rat CCD fragment, performed on rat genomic DNA (samples 1 and 4 as above) with primers *CCDP2 +CCDP3*. The abundant 1.5kb product was subsequently purified and inserted into a cloning vector.

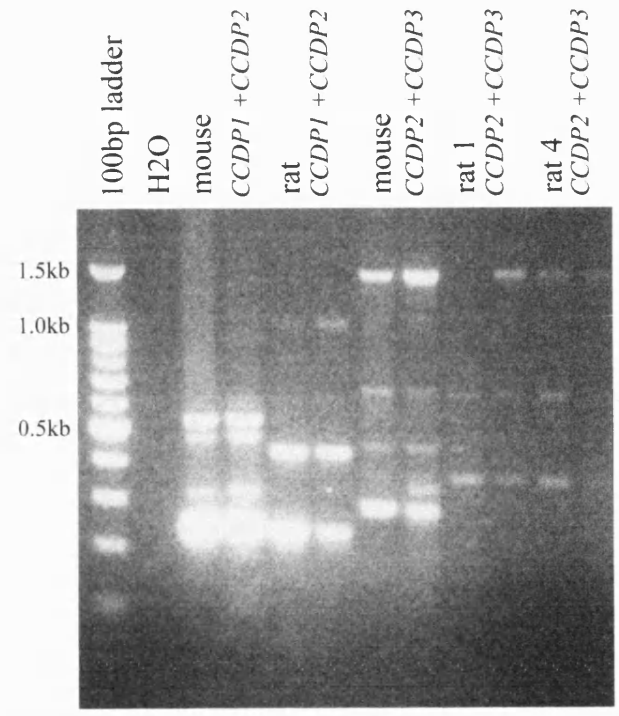


Figure 30a

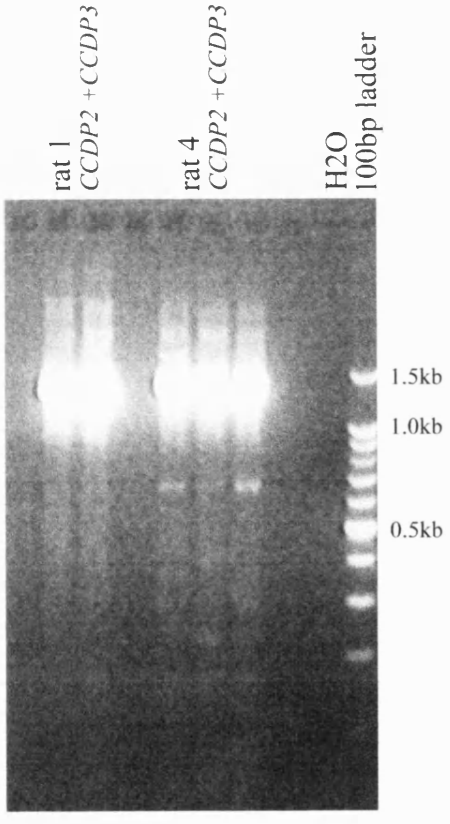


Figure 30b

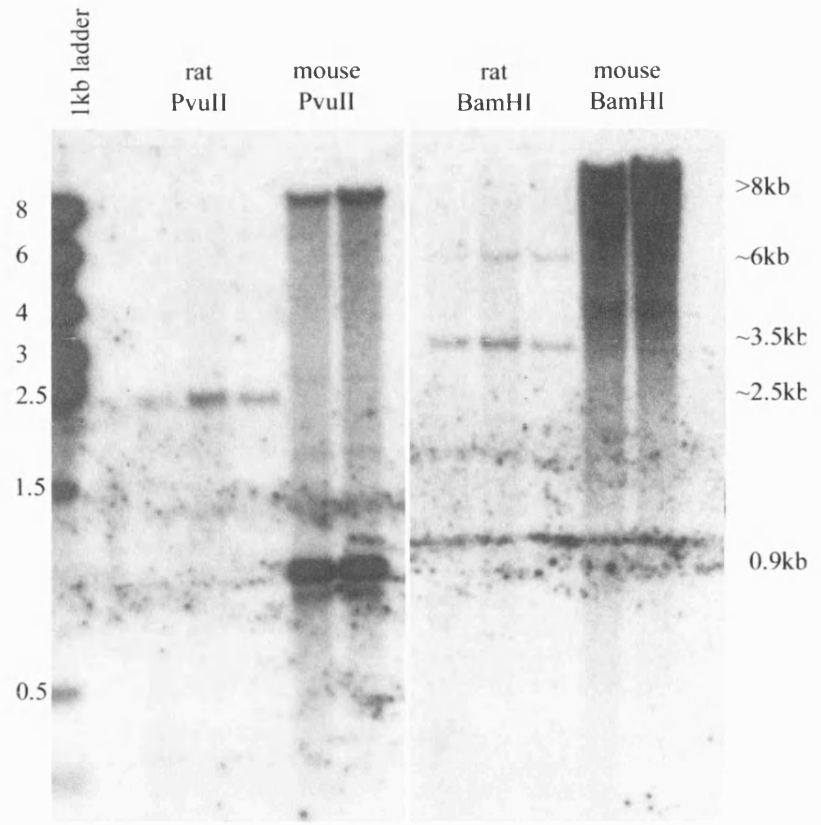


Figure 31. Southern blot showing restriction fragment length polymorphisms between the rat and mouse CCD region

Rat and mouse genomic DNA was digested with PvuII and BamHI and probed with the mouse CCD region. PvuII digestion of mouse genomic DNA reveals a major fragment of 0.9kb, and a fragment of >8kb, as predicted from the sequence data (see Table 18). The rat sequence lacks a second PvuII site, giving rise to an approximate 2.5kb fragment. The mouse CCD region contains no BamHI sites, and therefore the large fragments correspond to digestion outside the CCD region. The rat CCD is predicted to contain an internal BamHI site, giving rise to fragment sizes of approximately 3.5 and 6kb.

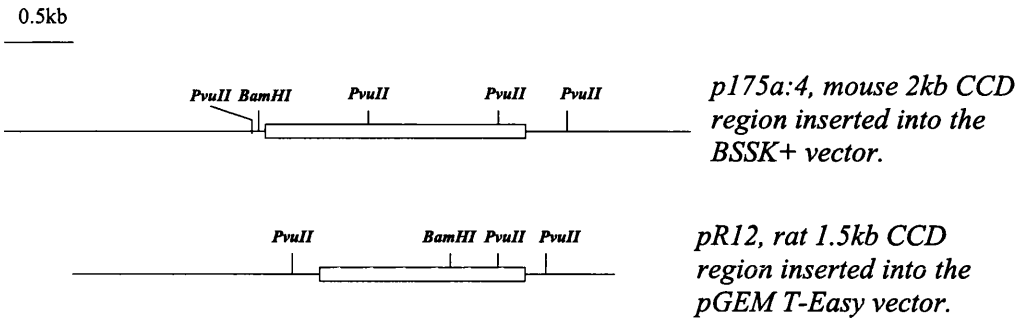


Figure 32. Schematic of the mouse and rat CCD regions in the context of their respective cloning vectors. The position of the restriction enzyme sites is based on sequencing of these regions. The rat and mouse CCD regions differ with respect to the position of several restriction sites. Polymorphic *BamHI* and *PvuII* sites are shown above. The predicted polymorphism at the sequence level compares favourably to an analysis of the same restriction sites in genomic DNA samples, examined by Southern blotting (see ***Figure 31***). The open rectangles represent CCD sequences whereas the line represents vector sequence.

	plasmid		***genomic***	
	PvuII	***BamHI***	***PvuII***	***BamHI***
mouse	0.5 0.9 0.9 2.5	4.9	0.9 >1	>2
rat	0.35 1.5 2.6	4.5	>1.4	>1.4

Table 18. Predicted restriction fragment sizes of mouse and rat plasmid and genomic DNA based on sequencing information. All sizes are rounded to the nearest 0.05kb.

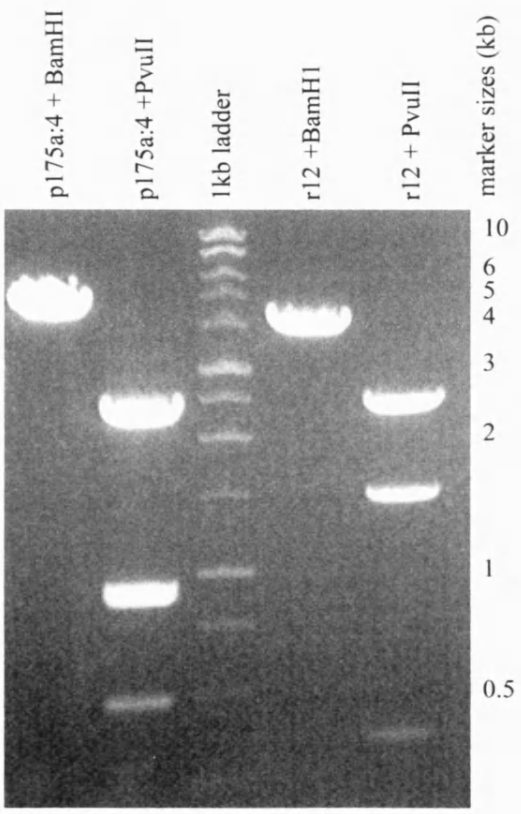


Figure 33. Restriction fragment length polymorphisms between the rat and mouse CCD region.

Mouse (p175a:4) and rat (r12) plasmid DNA was digested with PvuII and BamHI. p175a:4 is digested at a single site within the vector sequence of the plasmid, giving rise to a single fragment length of 4.9kb. p175a:4 is digested at four sites by PvuII, giving rise to fragment sizes of 0.5kb, 2.5kb, and a doublet at 0.9kb. r12 is cut once by BamHI within the rat CCD sequence, fiving rise to a fragment of 4.5kb. PvuII digests r12 at three sites, giving rise to fragment sizes of 2.6kb, 1.5kb and 0.35kb. These fragment sizes correspond to those predicted from the sequencing data, and are summarised in Figure 32 and Table 18

Region analysed		*Human*	*Rat*
Overall	*Length of sequence compared*	1615bp	1440bp
	% sequence identity to mouse	55%	83%
	number of gaps / %	306 / 19%	71 / 4%
Region 1	*Length of sequence compared*	118bp	93bp
	% sequence identity to mouse	83%	86%
	number of gaps / %	no gaps	2 / 2%
Region 2	*Length of sequence compared*	251bp	251bp
	% sequence identity to mouse	87%	97%
	number of gaps / %	2 / <1%	no gaps

Table 19. Summary of the data obtained from multiple species analysis of the CCD sequence. The sources of the mouse and rat sequences are detailed above. The human sequence was obtained from the NCBI database, accession number AC004556. Alignments, and the resulting percentage sequence identities, and gap scores were obtained using the 'BLAST 2 Sequences' algorithm at NCBI. Region 1 extends from 425bp-542bp of the mouse sequence (relative to the most *Igf2*-proximal nucleotide of the 2kb *EcoRI* clone). Region 2 extends from 1629bp-1879bp of the mouse sequence. The two regions were initially delineated as regions of significant homology to the human 11p15.5 region when a BLAST search was conducted with the mouse CCD nucleotide sequence, against the NCBI nr database. The sequence of the CCD from all three species, and their alignment performed using the WPI package is shown in ***Appendix 2***.

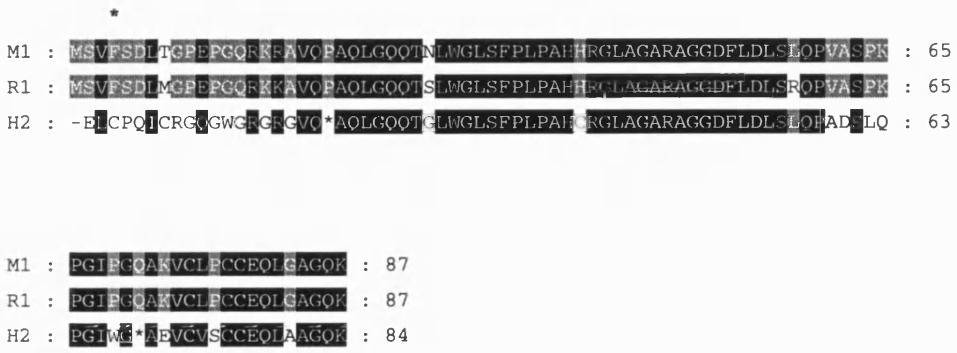


Figure 34. The amino acid sequence of a putative protein encoded by the ORF at Region 2 of the CCD. The mouse (M1), rat (R1) and human (H2) were translated from the position in the nucleotide sequence corresponding to a conserved ATG in the mouse and rat sequence. The asterisk above the sequence indicates where the consensus splice acceptor site present in all three species would dictate the beginning of the exon. The human sequence is frame shifted with respect to the mouse and rat sequence, due to the presence of two gaps in the alignment at nucleotides 1968bp and 1973bp of the human sequence (see ***Appendix 2***). This frame-shift also accounts for the poor conservation at the amino-acid level between the human and mouse/rat sequences in the beginning of this putative protein. The mouse and rat putative proteins share 84 amino acids of 87, a level of sequence identity of 97%. Asterisks within the amino acid sequence denote the position of stop codons.

Conclusions

The aim of the work discussed in this chapter was to answer the following questions: i) Does the *Igf2/H19* locus in species other than the mouse contain a region homologous to the CCD? ii) Are specific sequence motifs within the CCD region common to a range of species, and if so, what can the nature of such motifs tell us of CCD function?

The full sequence of a 2kb fragment containing the mouse CCD region was obtained and used to search a database (using the NCBI's sequence similarity search tool BLAST). The only hit from this search that displayed a significant level of similarity to the mouse CCD was a PAC genomic clone corresponding to human chromosome 11, region p15.5, which contains the *H19* gene and surrounding sequences. The mouse CCD region displayed high sequence identity to a region of this PAC at approximately 30kb upstream of the *H19* gene. The CCD is known to reside 32kb upstream of the *H19* gene in the mouse [151]. The human *Igf2/H19* locus therefore contains a region, which is homologous to the mouse CCD, in the same relative location. The commonality of location between the mouse and human CCD regions suggests that the sequences homologous to the CCD identified by cross-hybridisation analysis [151] in further mammalian species will also be found in the intergenic region between *Igf2* and *H19*. The conservation of its location in at least two mammalian species strengthens the proposal that the CCD contains an element that is important in the mechanism of regulation of *Igf2* and/or *H19*. It would be interesting to discover whether CCD-like sequences are also present in non-mammalian organisms, for example in chickens, where *Igf2* is known to be biallelically expressed [233].

The alignment of human and mouse CCD sequences revealed that the majority of the sequence identity was confined to two short regions. Region 1 lies at the most *Igf2*-proximal end of the CCD, and extends over approximately 120 nucleotides. Human and mouse Region 1 sequences share 83% sequence identity. The rat CCD sequence isolated by this study shares 86% sequence identity with the mouse in Region 1 (see ***Table 18***). The DNaseI hypersensitive sites demonstrated by Koide et. al. [151] to lie within the CCD have been mapped to the mouse CCD at nucleotides 432-460 and 512-541 (Dr J. F-X. Ainscough, pers. comm., see ***Figure 29***), both of which lie within Region 1.
Region 1 bears several characteristics of a regulatory region. Firstly, nuclease hypersensitive sites are thought to mark sequences involved in gene regulation (as

discussed in the **Introduction**). Secondly, Region 1 extends over a relatively short region, which suggests that sequence blocks such as those required for protein binding are conserved. A comparison of the region containing *H19* enhancer 2 between mouse and human (Accession numbers: mouse M23358¦MUSH19ENII, human AF091107), using the same search parameters as those used to compare CCD sequences (not shown), reveals a similar pattern of sequence conservation. The human and mouse enhancer regions share 83% sequence identity over 81 nucleotides. The mouse *H19* enhancer region has also been shown to be nuclease hypersensitive [151].

The delineation of a 120bp motif at Region 1 of the CCD by comparative sequencing will allow further investigation of this element. The use of techniques such as DNA-protein mobility assays with nuclear extracts can provide information about *trans*-acting factors that may interact with this region to bring about downstream events. The identification of such proteins could provide clues as to the role of the CCD region in the regulation of *Igf2/H19* expression and/or imprinting.

The role of Region 2 of the CCD is less clear. Region 2 extends over approximately 250 nucleotides, at the most *H19*-proximal end of the CCD (see ***Figure 29***). Comparisons between mouse, rat and human sequences reveal that this region is very highly conserved (mouse: human, 87%; mouse: rat, 97%). The extended homology throughout this region between mouse and human sequences suggested that it might contain a coding region. In both species, Region 2 contains an open reading frame that almost spans the entirety of the region.

The rat CCD sequence was obtained by PCR amplification using degenerate primers designed at two parts of the CCD that are highly conserved between mouse and human (in the middle of Region 1 and at the *H19*-proximal end of Region 2). The purpose of obtaining the sequence of the rat CCD was to enable a comparison of the rate of protein evolution at the ORF to a database of rates of evolution between 422 mouse-rat orthologs compiled by McVean and Hurst [68]. If the ORF at Region 2 is a coding region, its rate of protein evolution between mouse and rat would be predicted to lie within the distribution of the values obtained for the 422 orthologues.

The rat CCD, as well as the mouse and human contains a long ORF at Region 2. The mouse and rat sequences also share a conserved ATG at the 5' end of the ORF. Using this ATG as a start site, the mouse and rat sequences were translated *in silico* into a protein sequence, 87 amino acids in length. The rate of protein evolution and silent site evolution

was calculated, and compared to the mouse/rat database by L. Hurst (University of Bath). The putative coding region was found to have a rate of protein evolution at the low end of the distribution of the 422 genes. The rate of evolution is what would be expected of a protein coding sequence under stabilising selection. A complication of the analysis was that the silent sites too were found to have a low rate of evolution. There are several possible explanations of this finding. Firstly, Region 2 is not a coding region, and the conservation of all nucleotides is therefore is independent of their proposed codon position. Selection is instead acting on, e.g. nuclear factor binding sites. Secondly, the CCD may encode a transcript that is not translated, thus selection would be acting to conserve the secondary structure of the RNA, and therefore all nucleotides are equally likely to be conserved. Thirdly, Region 2 may encode a protein with a low rate of evolution at silent sites, as has previously been reported for imprinted genes [232]. Indeed, the rate of synonymous substitution of the putative gene is very similar to that of the neighbouring *Igf2* gene.

While the conserved ATG between rat and mouse was used as an initiation site for the translation of the nucleotide sequence, Region 2 may instead encode an internal exon. The CCD-internal end of Region 2 is predicted to act a splice acceptor site in all three species examined in this study. If Region 2 is but part of a gene, further sequences would be predicted to be conserved between species, which correspond to additional exons. Koide et. al. performed cross-hybridisation analysis on a series of non-overlapping clones that spanned the region between the *Igf2* and *H19* genes [151]. As well as demonstrating that the CCD region from the mouse was able to hybridise to genomic DNA in a number of mammalian species, this study also demonstrated that a second region in the mouse (the A22 clone) was able to hybridise to human genomic DNA. However, this region lies approximately 15kb upstream of the *H19* gene (i.e. downstream of the Region 2 putative exon), and therefore would not be predicted to encode the initiating exon for the putative gene, based upon the splice-site consensus data. No conserved regions were found upstream of the predicted Region 2 exon.

Sequence comparisons of mouse, rat and human Region 2 have been unable to unravel the function of this region, but raise a strong possibility that the CCD contains a coding region. Indirect evidence for the presence of a gene in this region must be supported by direct evidence provided by the identification of transcripts. Probes designed to detect the presence of transcripts from the ORF at Region 2 have been used in Northern hybridisation

and ribonuclease protection experiments to analyse RNA from a variety of samples of mouse tissues. To date, no transcripts originating from Region 2 have been identified (this study, and Koide et. al.[151]). A complication of such an analysis is that the putative gene could be active in a very narrow time window of development, in a small subset of tissues, or transcribed at very low abundance, making it difficult to equivocally rule out the existence of a gene at the CCD. Further work could be to screen cDNA libraries (which provide good coverage of expressed transcripts) for clones homologous to the ORF; or by a program of analysis of RNA by Northern hybridisation, ribonuclease protection or Reverse-Transcription PCR, with an aim to cover the widest possible range of tissues and developmental stages.

In summary, comparative sequencing of the CCD between mouse, rat and human has revealed that the location of the CCD is conserved (at least between mouse and human). Furthermore, two highly conserved regions within the CCD have been identified, one of which displays several features of a regulatory region (Region 1). Region 2 contains an ORF that is highly conserved between the three species at both the nucleotide and amino-acid level. However, no direct evidence of a transcript originating from this region has been obtained.

The comparative sequencing approach has provided a valuable tool in isolating highly conserved regions within the CCD that may well correspond to the minimal sequences required for its function. The identification of these minimal sequences will aid further analysis of possible *trans*-interactions between nucleotide sequences and the effector proteins that are required for the function of this element.

CHAPTER 9: CONCLUSIONS

Tissue specific expression

As was discussed in the **Introduction**, the spatial and temporal expression patterns of the *Igf2* and *H19* genes is complex, and relies upon interactions between the promoters of these genes with multiple enhancer and silencer elements. The similarity in the expression patterns of the two genes implies that they share at least some of these regulatory elements. Indeed, in the case of expression in tissues with an endodermal component (e.g. liver and gut), both genes have been shown to be regulated by a common set of enhancers [104].

The developmental timing by which the two genes become monoallelically expressed differs, and is tissue specific. For example, *Igf2* is expressed with a paternal bias from the earliest stages of its expression [86], and biallelic expression of this gene is limited to the exchange tissues of the brain [70]. *H19* is expressed equally from both parental alleles during early developmental stages [86], and shows biallelic expression in some placental cell types [116], but is imprinted in the exchange tissues of the brain [98]. In the case of both genes, their imprinting status is dependent upon tissue type, and therefore the mechanisms that confer tissue-specific expression might also play a role in the imprinting mechanism.

The CCD was initially characterised by Koide et. al. [151] as a region approximately midway between *Igf2* and *H19* that displayed enhanced sensitivity to the nuclease DNase I, and was conserved in a number of mammalian species. The CCD provides a strong candidate for an additional enhancer or imprinting control element at this locus. Both the *Igf2* and *H19* genes are expressed at high levels during development in mesodermal tissues [95, 78]. An initial hypothesis during the course of this study was that the CCD might harbour an enhancer of expression in tissues of mesodermal origin. A 130kb YAC transgene that contained sequences from P1 of the *Igf2* gene to approximately 35kb downstream of the *H19* gene [36] demonstrated the appropriate expression of both an *Igf2-LacZ* fusion gene, and *H19* in several mesodermal tissues, including skeletal muscle and the tongue. Sequences responsible for gene expression in these tissues therefore lie in the immediate vicinity of the two genes. Three strains of transgenic mice bearing a construct containing the CCD *in-cis* to a P3-*luciferase* reporter (the ***H*** construct, see ***Figure 3*** in **Chapter 1**) were examined for reporter gene expression in the tongue and skeletal muscle of neonatal mice. Very low levels of *luciferase* gene activity were observed in these

tissues, demonstrating that the CCD is not an enhancer of P3 in these mesodermal tissues. Appropriate expression patterns were observed from transgenic lines constructed with the ***A*** construct (see ***Figure 3***). This construct contains the reporter gene *in-cis* to the *H19* enhancers, and in this system these enhancers upregulated gene expression from P3 in liver and kidney, but not in tongue and skeletal muscle. To test for any possible interactions between the CCD and the *H19* enhancers, gene expression levels were quantitated in tissues from a transgenic line bearing both elements (***Q*** lines, see ***Figure 3***), and compared to mean expression levels in the same tissues from ***A*** or ***H*** lines. In three independent ***Q*** lines, no new expression patterns were observed in any of the tissues analysed, and expression levels were found to be quantitatively similar between ***Q*** and ***A*** lines in the endodermal tissues, and between all three lines in the mesodermal tissues. The possibility of synergy between the CCD and the *H19* enhancers was tested at an additional time point, e14.5. In the bodies of transgenic mice at e14.5, the median level of gene expression from the three ***Q*** lines was slightly elevated over the median expression level from ***A*** lines. However, due to the variation in expression levels between transgenic lines bearing a particular construct, and the small number of lines tested, the significance of this result is unclear.

In agreement with a previous report [108], the CCD was shown to drive reporter gene expression in the brains of transgenic mice at D1. The major sites of *Igf2* and *H19* expression in the brain are the choroid plexus and leptomeninges [97, 98]. Gene expression from reporter constructs was found to be localised to the exchange tissues, and was not then thought to be a result of ectopic gene expression. In addition to the CCD, the *H19* enhancers were also found to have some activity in the exchange tissues, in agreement with a previous report of expression from these enhancers in the meninges but not the choroid plexus [109].

In parallel with the *in vivo* study of reporter gene activity in the tissues of transgenic mice, a series of *in vitro* assays were performed. A number of immortal cell lines were transiently transfected with the same reporter constructs as used to create the transgenic lines (***M***, ***H***, ***A*** and ***Q***, see ***Figure 3***, **Chapter 1**), in order to determine the effects of the CCD and the *H19* enhancers on the levels of reporter gene activity from the P3 promoter. The cell lines were chosen to be representative of those tissues analysed *in-vivo*, i.e., the hepatocyte cell line *Hep3B*, the kidney line, *Cos7*, and the skeletal muscle cell line *C2*. The relative levels of expression from reporter constructs *in vitro* largely

reflected the expression patterns of the constructs *in vivo*. As has been shown previously, the *H19* enhancers elevate the levels of reporter gene expression above basal promoter levels in *Hep3B* cells [103]. In all other cell lines examined, the *H19* enhancers did not drive reporter gene activity over basal levels observed with P3 alone. This result confirms the tissue specificity of the enhancers shown in this study, and previously [103]. The CCD did not upregulate P3 in any cell line examined, again in agreement with the *in vivo* analysis of transgene expression in liver, skeletal muscle and kidney, and confirming that the CCD does not contain an enhancer for these tissues.

As the CCD appeared to be active in a limited number of cell types, expression of reporter constructs from one of these cell types, the choroid plexus, was analysed in an *in vitro* system. This necessitated the optimisation of a new technique for the creation of primary cultures of choroid plexus tissue. With several modifications, an existing protocol for the primary culture of porcine choroid plexus was adapted for the culture of neonatal mouse tissue. The identity of the cultured cells was confirmed as choroid plexus epithelium by analysis of cell morphology and gene expression. Importantly, *Igf2* was shown to be expressed in these cells. However, due to technical problems with transient transfection of the primary cultures, differences in the levels of gene expression between reporter constructs could not be confirmed. Verification of CCD-mediated expression in an *in vitro* choroid plexus culture system awaits further work.

A second purpose of developing an *in vitro* model in which CCD activity could be demonstrated was in order to isolate minimal functional sequences within this region with activity. Once minimal sequences had been identified, they could be used to perform DNA-protein binding assays to characterise *trans*- acting proteins. The CCD did not affect reporter gene activity in any *in vitro* system tested, so this analysis could not be carried out.

Another method by which to localise sequences that may have a regulatory function is by searching for conserved regions between different species. Genomic DNA with homology to the CCD was shown to exist in a number of mammalian species by cross-hybridisation analysis [151]. In this study, the full sequence of the mouse CCD was obtained, and used to screen a nucleotide database for similar sequences in other organisms. The only significant hit from this search was found to a region of human chromosome 11p15.5, approximately 30kb upstream of the *H19* gene. The human *Igf2/H19* domain thus contains an element homologous to the CCD. The sequence identity between the CCDs from mouse and human largely resides in two regions. Region 1

comprises approximately 120 nucleotides, and contains the DNaseI hypersensitive sites [151]. Region 2 spans approximately 250 nucleotides, and bears several characteristics of a coding region. Region 2 contains a long open reading frame in mouse, rat and human, and is flanked at the 5' end with a 40 nucleotide sequence that could act in all three organisms as a splice acceptor site. The rate of divergence of this sequence between rat and mouse is compatible with that of a coding region under stabilising selection. However, no transcripts originating from this region have been identified in this study, or in others [151, 229]. The CCD thus potentially contains at least two distinct regions, which may represent regulatory and/or coding activity.

Recently [229], the role of Region 1 of the CCD was investigated by the analysis of the effect of its deletion in the context of a 130kb *Igf2/H19* YAC transgene. At low copy number an *Igf2-LacZ* fusion construct displays expression in a wide subset of tissues known to express *Igf2*. The YAC also contains a copy of the *H19* gene, and both genes are imprinted appropriately. Deletion of a 1kb region containing CCD Region 1 (in this study, the most 5' 1kb of the CCD clone) was carried out by *Cre*-mediated recombination between flanking *loxP* sites. The expression of both genes from the deleted YAC was indistinguishable from the parent construct until e17. At this stage, the previously silent *Igf2-LacZ* fusion gene was activated on the maternal allele in a highly specific subset of tissues, including the intrinsic muscles of the tongue, the craniofacial muscles, as well as other skeletal muscles throughout the body. *H19* gene expression from the YAC was not effected. The authors concluded that this region contains a silencer that limits the expression of the *Igf2* gene in late embryogenesis in skeletal muscle.

The deleted YAC construct continued to express the *LacZ* fusion gene in the choroid plexus and meninges of the bran. This suggests that the brain enhancer activity of the CCD demonstrated in this study does not reside in the first 1kb of the CCD that contains Region 1. The second half of the CCD must contain the enhancer activity, with Region 2 providing a good candidate.

In another recent study, several additional candidate regulatory elements for *Igf2* and *H19* have been identified by a comparative sequencing approach [234]. This study determined and compared the sequences of the human and mouse *H19* region over 41kb (extending from 8kb upstream to 30kb downstream of the *H19* gene). Ten evolutionarily conserved segments were identified; two of which were co-incident with the known endodermal enhancers. The remaining eight were inserted into *LacZ* reporter constructs,

and used to create transgenic mice. Reporter assays showed that five of the eight elements functioned as enhancers in specific mesodermal and ectodermal tissues (including the limb buds, myotome, rib primordia and intercostal muscles, at e12.5). These elements would all be predicted to lie on the 130kb YAC transgene [36], and presumably are responsible for the temporal and spatial expression patterns of *Igf2-LacZ* and *H19* from that transgene. It would appear then that the majority of the embryonic enhancers for *Igf2* and *H19* lie downstream of the *H19* gene, though the action of these enhancer elements on their target genes can only be confirmed by the analysis of the effects of deleting these elements from their normal genomic location. The ten conserved elements characterised in the comparative study did not include an enhancer for the exchange tissues of the brain. The CCD region thus provides the only *cis* element identified to date that can drive gene expression in the choroid plexus.

Parent of origin-specific effects

A previous investigation of the mode of expression of the P3-*luciferase* transgenes detailed in this report demonstrated that the *H19* enhancers regularly conferred a parental bias on the levels of gene expression from the reporter construct. In the presence of the *H19* enhancers, *luciferase* expression levels were found to be 2-3 fold higher following maternal transmission of the transgene, as compared to levels following paternal transmission of the transgene [108]. In this report, gene expression levels were compared following maternal versus paternal transmission of ***A***, ***H*** and ***Q*** transgenes, to determine if the phenomenon was repeatable. In addition, since the CCD might act to override the imprinting signal, the analysis of ***Q*** lines was to test for an interaction between the *H19* enhancers and the CCD. A comparison of three transgenic lines bearing each construct demonstrated that the *H19* enhancers regularly imposed a maternal bias on gene expression, in both the presence and the absence of the CCD, at two different time points. However, a strong paternal bias was seen in some lines in some tissues. ***H*** lines never conferred reproducible parental origin-specific effects at a high level of statistical significance, demonstrating that this element cannot mediate imprinting effects alone in the tissues analysed here. As in the previous study [108], the maternal bias was generally modest (2-5 fold). In addition, the levels of reporter gene activity were found to be largely independent of gross levels of reporter gene and promoter methylation. The *H19* enhancer region was also examined for methylation differences following maternal or paternal

transmission of the transgene, but the methylation status of only a single CpG dinucleotide was detectable by the assay. As with the *luciferase* reporter gene and promoter region, no gross changes in methylation were observed between samples.

The mechanism of imprinting of the *Igf2* and *H19* genes appears to be linked to the tissue specificity of these genes. Expression of both genes can be blocked by the action of tissue-specific silencers [176, 229, 164]. In addition, the endodermal expression of *Igf2* is prevented by the action of a methylation-sensitive boundary element that prevents the access of the *Igf2* promoters to distal enhancers [170, 169, 155]. The 'imprint' that leads to inactivation of the appropriate parental allele of these genes does not reside in a single element, and instead is the consequence of the action of multiple sequences that act in different tissues, possibly through direct interaction with tissue-specific enhancer elements. The modest imprinting effect observed with the *H19* enhancers suggests that enhancer elements themselves play a direct role in the imprinting process. The *H19* enhancers have previously been implicated in the mechanism of imprinting. In a study of *H19* transgenes [37] containing an internally deleted *H19* gene region, 4kb of 5' and 8kb of 3' flanking DNA was found to be necessary to maintain the imprinting status of the reporter gene. If the 3' flanking region containing the enhancers was not present in the transgene construct, the reporter gene was not expressed, and in addition, was not appropriately methylated at the 5' flanking region.

How does *Igf2* escape imprinting in the choroid plexus and leptomeninges? The identification of the CCD as an enhancer of *Igf2* in these tissues may lead to the understanding of how *Igf2* achieves biallelic gene expression. It appears that the majority of enhancer elements for *Igf2* may lie on the opposite side (from *Igf2*) of a boundary element [103, 234, 170, 169, 155]. The CCD lies between this boundary and *Igf2*, so presumably is not affected by it. Indeed, when the *H19* enhancers were relocated to the intergenic region, approximately 50kb upstream of the *H19* gene, *Igf2* expression became biallelic in the tissues in which this element is active [163]. The CCD may therefore confer biallelic expression to *Igf2* as a result of its position. However, the differentially methylated regions of the *Igf2* gene (DMR1 and DMR2, see ***Figure 1*, Chapter 1**) are methylated on both parental alleles in the choroid plexus [7], suggesting that an additional epigenetic mechanism may also be involved.

Further work

A full investigation of the role of the CCD has been prevented by the failure to establish *in-situ* patterns of reporter genes in the presence of the CCD. The strategy of immunohistochemical staining for *luciferase* protein was employed to gain this information, but was not successful. The expression patterns of reporter genes from existing *luciferase* transgenic lines could be visualised by the technique of *in-situ* hybridisation for *luciferase* mRNA. This technique was not adopted during the course of this investigation for the reasons discussed in **Chapter 3,** and instead work was begun to construct a new series of transgenic lines utilising the reporter gene *LacZ.* A construct was created that contained the *LacZ* gene driven from P3, in the presence of the CCD (*C* lines, see ***Figure 3***). This construct was microinjected into 1-cell embryos, and the injected embryos implanted into the uterus of a pseudopregnant foster mother. To date, two offspring have been recovered in which the transgene successfully integrated, at least one of which transmits the transgene through the germline (data not shown). Time constraints have prevented any further analysis of these transgenics, but in future it will be interesting to visualise the action of the CCD on gene expression at the single cell level.

The primary cultures of choroid plexus epithelium developed in this investigation provide a resource in which to study *Igf2* expression in the choroid plexus, as well as determining how this gene escapes imprinting in this tissue. The parental alleles of *Igf2* can be identified by culturing cells from the offspring of a cross between parents that are polymorphic for a marker within the *Igf2* gene (for example, a restriction fragment length polymorphism between two mouse strains). As discussed above, the methylation status of *Igf2* at DMR1 and DMR2 is unusual in the choroid plexus, in that both alleles display the paternal epigenotype [7]. Whether this is a cause or consequence of the biallelic mode of *Igf2* expression can be tested by treating choroid plexus cells, with distinguishable parental alleles, with the DNA demethylating agent 5-aza-2' –deoxycytidine. In a tumour cell line that displayed loss of imprinting of *Igf2*, this state was reverted to monoallelic expression by the addition of 5-aza-2' –deoxycytidine to the culture medium [235]. In addition, chromatin modifications such as histone acetylation may play a role in the biallelic expression of *Igf2*. The histone acetylation status of DNA can be manipulated in a cell culture system by the addition of the histone deacetylase inhibitor Trichostatin A [116].

The creation of an immortal cell line from choroid plexus cultures would greatly facilitate the execution of the experiments suggested above. There are a number of reasons

to suppose that choroid plexus cultures derived from mice lacking *p53* and *pRB* family function would be immortal. A cell line has recently been developed from primary cultures of rat choroid plexus, by treating the cultures with the immortalising *SV40 T-antigen* [236], which inactivates these two proteins [231]. Saenz Robles et. al. [231] have created a transgenic mouse line that express *T-antigen* in the choroid plexus epithelium (under the control of the *transthyretin* promoter). The creation of a cell line from the choroid plexus of these mice will have the advantage that the only immortal cells present in the population will be those of the choroid plexus epithelium, and therefore contaminating cells will be at a selective disadvantage.

The identification of an element that drives *Igf2* expression in the exchange tissues of the brain, and the creation of a protocol by which to isolate and culture choroid plexus cells from mouse tissue will allow further study of the mechanism by which *Igf2* escapes imprinting. These studies may in turn aid in the identification of the mechanisms by which *Igf2* becomes biallelically expressed in pathological circumstances.

APPENDIX 1

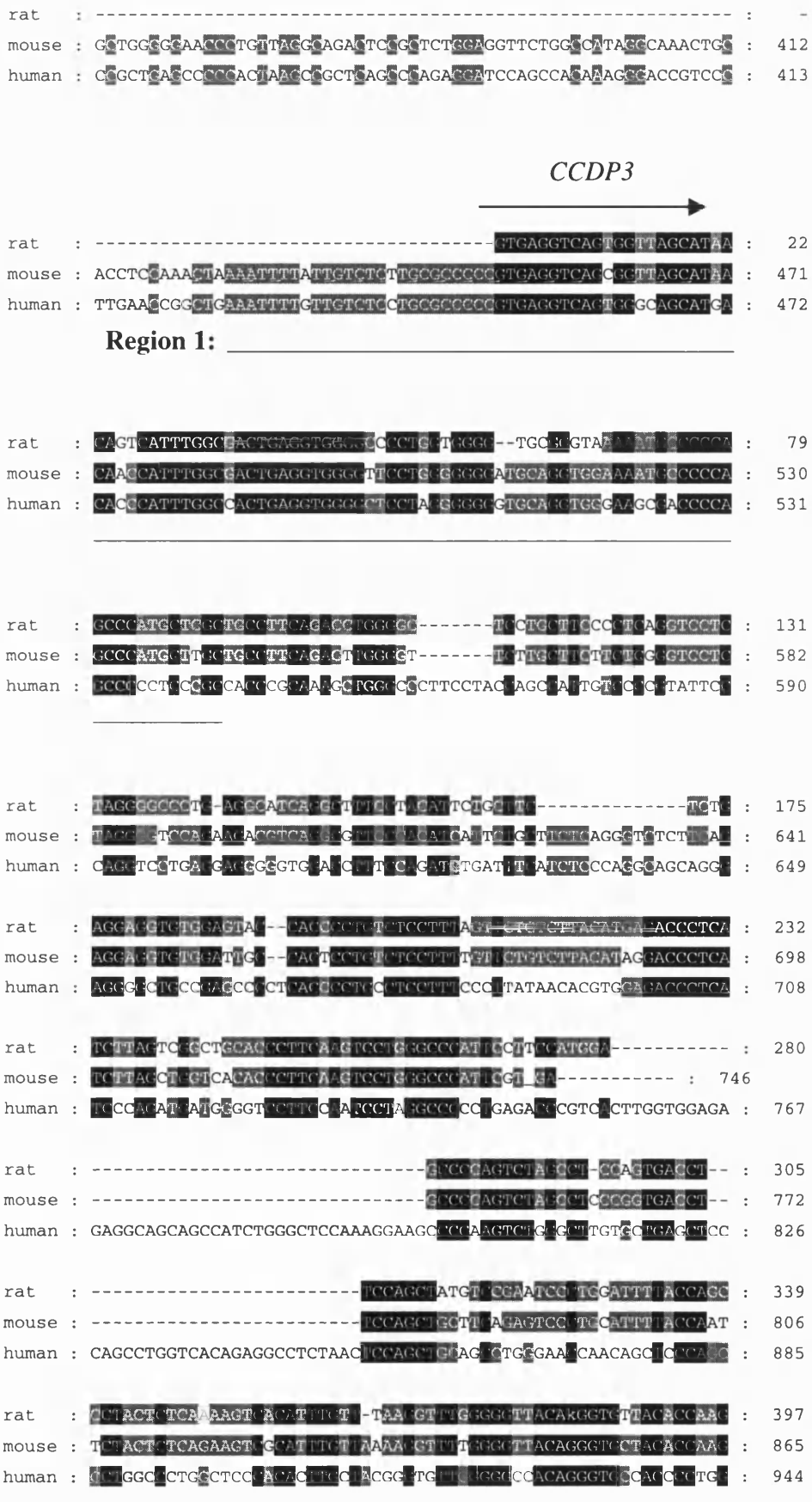
construct	line	copy number	Mean *luciferase* specific activity (*ng luciferase/mg protein*) following male transmission			transgene methylation status
			brain	tongue	liver	
A	*Ayah*	1-2	$2.36x10^{-3}$	$6.54x10^{-2}$	7.52	+++
	Archy	4-6	$1.52x10^{-2}$	$4.72x10^{-2}$	25.66	-
	Alicia	1-2	$1.21x10^{-2}$	$1.82x10^{-2}$	4.74	+++
	Axe	10	nd			nd
Q	*Quark*	2-4	$1.69x10^{-2}$	$2.10x10^{-4}$	2.68	-
	Quasar	1-2	0.72	$8.84x10^{-3}$	10.41	-
	Quiche	nd	0.32	$6.58x10^{-2}$	14.12	nd
H	*Hamish*	nd	$2.39x10^{-2}$	$3.20x10^{-4}$	$5.10x10^{-6}$	nd
	Harold	2	0.14	$9.86x10^{-2}$	$3.00x10^{-4}$	nd
	Holly	3	0.29	$6.80x10^{-3}$	$4.00x10^{-4}$	nd

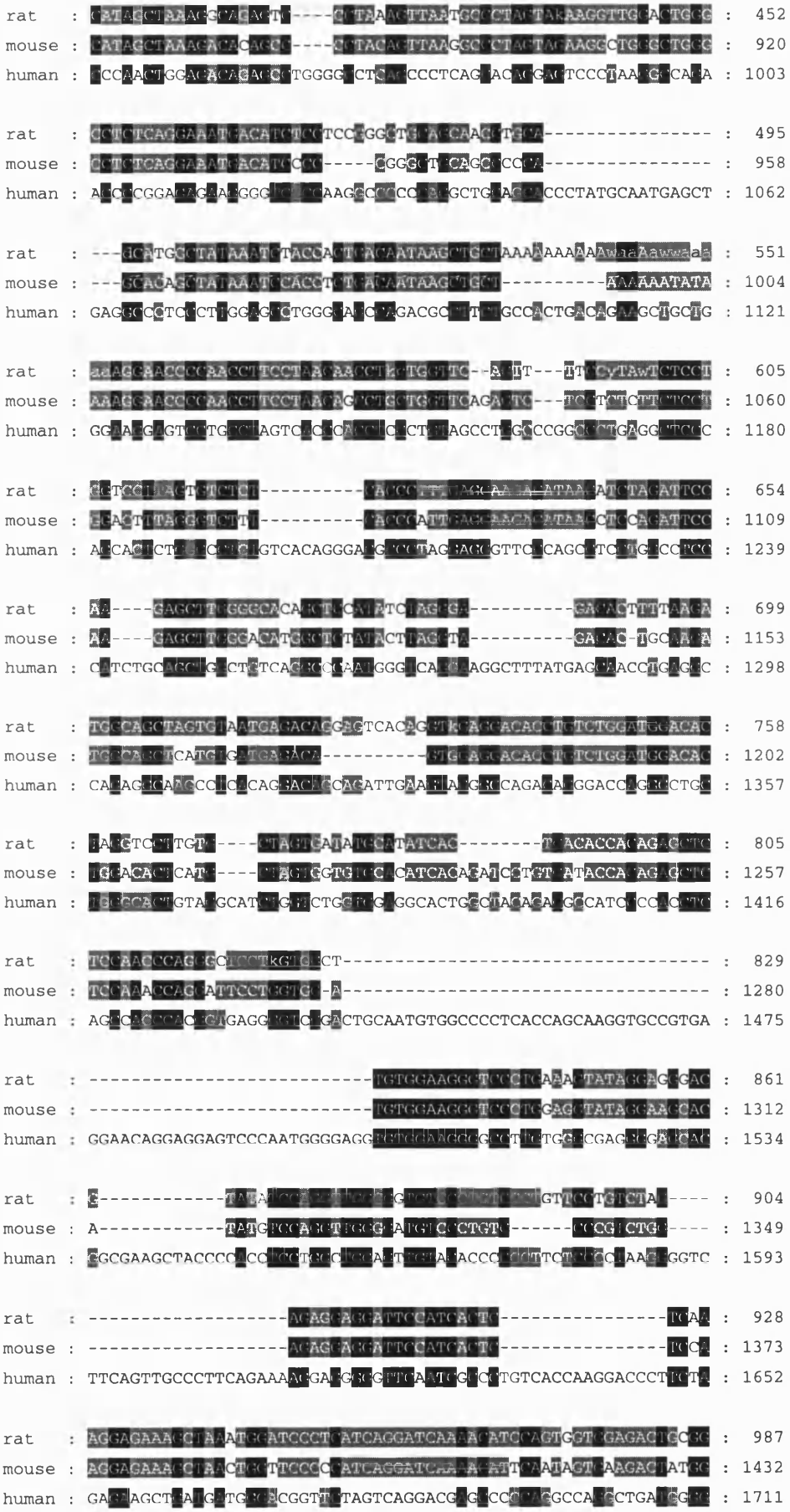
Appendix 1. Transgenic lines used in this study, correlated to copy number (determined by visual estimation of the signal on Southern blots, in comparison with a single-copy endogenous gene signal, [108] and data not shown), reporter gene specific activity, and methylation status (+++, highly methylated, -, not methylated). There appears to be no correlation between transgene copy number, gene expression levels, and the methylation status of the transgene.

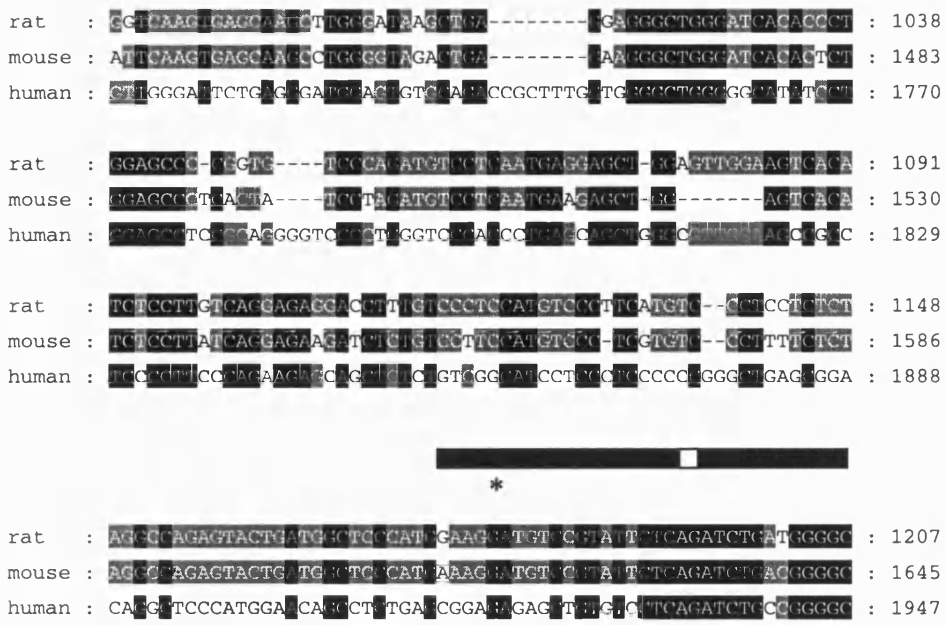
APPENDIX 2

Pileup of mouse, human and rat CCD sequences (see **Chapter 8**). Grey shading indicates agreement between two species, black shading indicates agreement between three species. The positions of PCR primers used to clone the rat CCD are shown as arrows above the sequence. Conserved sequence blocks discussed in the text as Region1 and Region2 are noted below the sequence. The ATG codon chosen as the start site for analysis of protein sequence homology between mouse and rat is indicated as an asterisk above the sequence. The position of a consensus splice acceptor site is indicated as a thick black line above the sequence, with the intron/exon boundary indicated as a white box. The origin of 1.47kb of rat CCD sequence and 2kb of mouse CCD sequence is described in **Chapter 8**, and the human CCD sequence was obtained from the NCBI database, accession number AC004556. The sequence displayed here represents the reverse complement of nucleotides 90481-92700. The alignment was performed using the Pileup algorithm of the Wisconsin Package Version 8.0, using the default parameter settings.

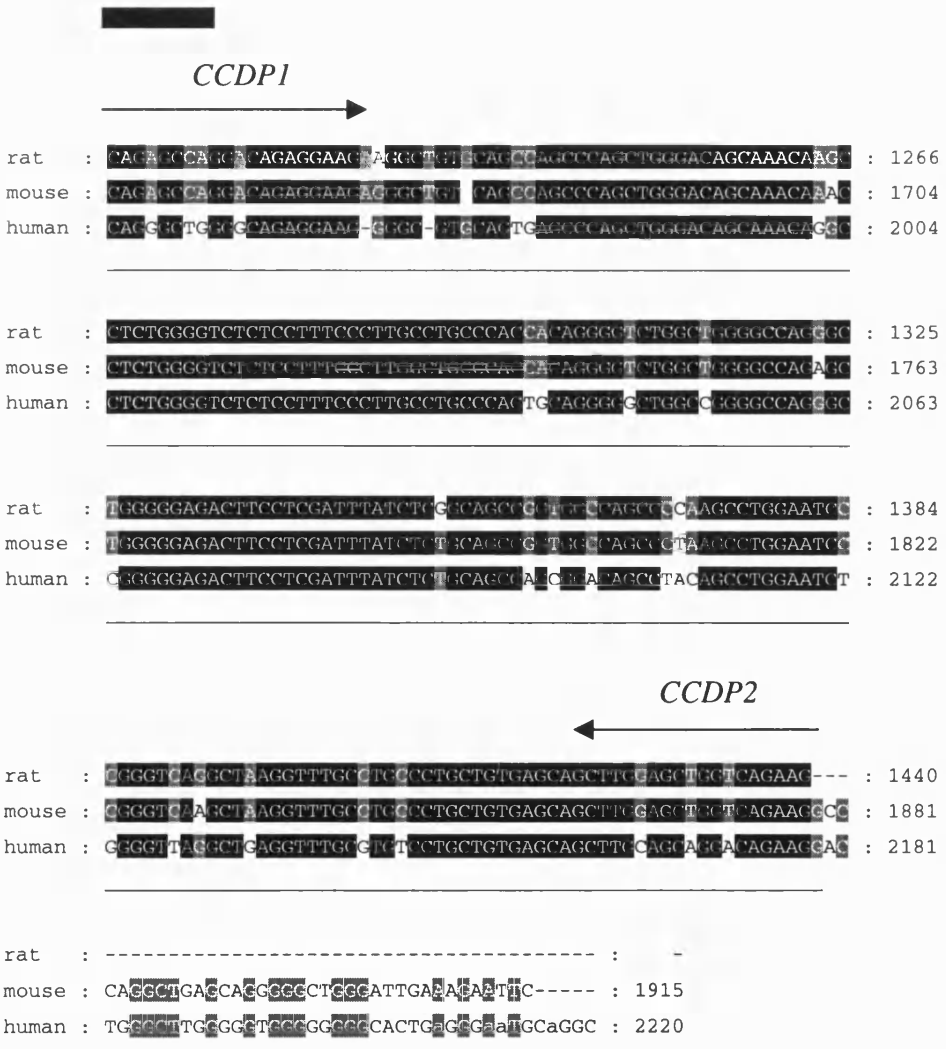
```
rat   : ------------------------------------------------------------ :    -
mouse : -GAATTCCAGCTACCATCCTCCGTCACAGCTGCTGATGGCAGCCCTCCAGCTCTGTAA :   58
human : CTTTGGAGTCCTCAGCCACCTCTCCCAAGCTCGTGGCGCAGCGCCCACAGCTGGCTCCCA :   59

rat   : ------------------------------------------------------------ :    -
mouse : CAGCAAAGAAAGGCTGGGCCTGCCTGTCTCTAGGGTGGCAGTGGGGGTTGGTCCCTGCC :  117
human : GGCCCTTCCCAGAGCTTCCGGCTCCCCTCGGCTGCCTTGGGACCTGTGGCCCCCTCCGCCC :  118

rat   : ------------------------------------------------------------ :    -
mouse : TACCCTAGTGCCATACAGAAAAGTATTCTTCTCAATGCCTTGCCCCTGAGAAGCAAGAA :  176
human : CAGGTGCAAAGCCCTTCCTCCCGCCCCCACCGCCCCCAGGAAGCCTCAGTCCCCAGCCT :  177

rat   : ------------------------------------------------------------ :    -
mouse : GGGGAAGCGGGGGTGGGGGTGGTCGATGGGGGTGGAGAGGAGGTGGGAGAGAGCCCTG :  235
human : GTCTGGGCTGCCCCTGCCCCAAGGAGAGGCAGTGCGGAAGTGGAGAGGAGAAGGGAGAG :  236

rat   : ------------------------------------------------------------ :    -
mouse : AGGAAGACAAGAGGAAGGGAAGCAGAAGGGAGTGGGCAGAAAGACGTTAGAGGCCACA :  294
human : CGGCCTGGGGAGCAGGACAGAGCCAAGAGGGGCTGCAAGGTCAGGGCACATTGTTCAGG :  295

rat   : ------------------------------------------------------------ :    -
mouse : GAACCGCCAGGGGTAGCATGCCTATCAGCAAGCTAGCCAGGAGTGGGCTCTCCAGCAGA :  353
human : GTGGGAGGTGAGTGGTGCTCAGGCAGGGTGGCTCGTGTTTGAGAGACGCCTGGGGAGCCC :  354
```


CCDP3
Region 1:





Region 2:



BIBLIOGRAPHY

1. Surani, M.A.H., Barton, S.C. & Norris, M.L. Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* **308**, 548-550 (1984).
2. McGrath, J. & Solter, D. Completion of Mouse Embryogenesis Requires Both the Maternal and Paternal Genomes. *Cell* **37**, 179-183 (1984).
3. Mann, J.R. & Lovell-Badge, R.H. Inviability of parthenogenones is determined by pronuclei, not egg cytoplasm. *Nature* **310**, 66-67 (1984).
4. Morison, I.M. & Reeve, A.E. A catalogue of imprinted genes and parent-of-origin effects in humans and animals. *Human Molecular Genetics* **7**, 1599-1609 (1998).
5. Caspary, T., Cleary, M.A., Baker, C.C., Guan, X.-J. & Tilghman, S.M. Multiple Mechanisms Regulate Imprinting of the Mouse Distal Chromosome 7 Cluster. *Molecular and Cellular Biology* **18**, 3466-3474 (1998).
6. Moore, T., Constancia, M., Zubair, M., Balleul, B., Feil, R., Sasaki, H. & Reik, W. Multiple imprinted sense and antisense transcripts, differential methylation, and tandem repeats in a putative imprinting control region upstream of mouse *Igf2*. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 12509-12514 (1997).
7. Feil, R., Walter, J., Allen, N.D. & Reik, W. Developmental control of allelic methylation in the imprinted mouse *Igf2*and *H19* genes. *Development* **120**, 2933-2943 (1994).
8. Sasaki, H., Jones, P.A., Chaillet, J.R., Ferguson-Smith, A.C., Barton, S.C., Reik, W. & Surani, M.A. Parental imprinting: potentially active chromatin of the repressed maternal allele of the mouse Insulin-like Growth Factor-II (Igf2) gene. *Genes & Development* **6**, 1843-1856 (1992).
9. Ferguson-Smith, A.C., Sasaki, H., Cattanach, B.M. & Surani, M.A. Parental-origin-specific epigenetic modification of the mouse H19 gene. *Nature* **362**, 751-755 (1993).
10. Stoger, R., Kubicka, P., Lui, C.-G., Kafri, T., Razin, A., Cedar, H. & Barlow, D.P. Maternal-Specific Methylation of the Imprinted Mouse *Igf2r* Locus Identifies the Expressed Locus as Carrying the Imprinting Signal. *Cell* **73**, 61-71 (1993).
11. Smilinich, N.J., Day, C.D., Fitzpatrick, G.V., Caldwell, G.M., Lossie, A.C., Cooper, P.R., Smallwood, A.C., Joyce, J.A., Schofield, P.N., Reik, W., Nicholls,

R.D., Weksberg, R., Driscoll, D.J., Mahler, E.R., Shows, T.B. & Higgins, M.J. A maternally methylated CpG island in *KvLQT1* is associated with an antisense paternal transcript and loss of imprinting in Beckwith-Wiedemann syndrome. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 8064-8069 (1999).

12. Schumacher, A., Buiting, K., Doefler, W. & Horsthemke, B. Methylation analysis of the PWS/AS region does not support an enhancer-competition model. *Nature Genetics* **19**, 324-325 (1998).
13. Wutz, A., Smrzka, O.W., Schweifer, N., Schellander, K., Wagner, E.F. & Barlow, D.P. Imprinted expression of the *Igf2r* gene depends on an intronic CpG island. *Nature* **389**, 745-749 (1997).
14. Rivkin, M., Rosen, K.M. & Villa-Komaroff, L. Identification of an Antisense Transcript From the IGF-II locus in Mouse. *Molecular Reproduction and Development* **35**, 394-397 (1993).
15. Mitsuya, K., Meguso, M., Lee, M.P., Katoh, M., Schulz, T.C., Kugoh, H., Yoshida, M.A., Nikawa, N., Feinburg, A.P. & Ohimura, M. *LIT1*, an imprinted antisense RNA in the human *KvLQT1* locus identified by screening for differentially expressed transcripts using monochromosomal hybrids. *Human Molecular Genetics* **8**, 1209-1217 (1999).
16. Rougeulle, C., Cardoso, C., Fontes, M., Colleaux, L. & Lalande, M. An imprinted antisense RNA overalps *UBE3A* and a second maternally expressed transcript. *Nature Genetics* **19**, 15-16 (1998).
17. Lee, J.T. & Lu, N. Targeted Mutagenesis of *Tsix* Leads to Nonrandom X inactivation. *Cell* **99**, 47-57 (1999).
18. Sleutels, F., Barlow, D.P. & Lyle, R. The uniqueness of the imprinting mechanism. *Current Opinion in Genetics and Development* **10**, 229-233 (2000).
19. Barlow, D.P. Competition-a common motif for the imprinting mechanism? *The EMBO Journal* **16**, 6899-6905 (1997).
20. Lyle, R., Watanabe, D., te Vruchte, D., Lerchner, W., W., S.O., Wutz, A., Schageman, J., Hahner, L., Davies, C. & Barlow, D.P. The imprinted antisense RNA at the *Igf2r* locus overlaps but does not imprint *Mas1*. *Nature Genetics* **25**, 19-21 (2000).
21. Brannan, C.I. & Bartolomei, M.S. Mechanisms of genomic imprinting. *Current Opinions in Genetics and Development* **9**, 164-170 (1999).

22. Mann, M.R.W. & Bartolomei, M.S. Maintaining imprinting. *Nature Genetics* **25**, 4-5 (2000).

23. Horsthemke, B. Imprinting in the Praeder-Willi/Angelman syndrome region on human chromosome 15. in *Genomic Imprinting* (eds. Reik, W. & Surani, A.) (Oxford University Press, Oxford, New York, Tokyo, 1997).

24. Bielinska, B., Blaydes, S.M., Buiting, K., Yang, T., Krajewska-Walsek, M., Horsthemke, B. & Brannan, C.I. *De novo* deletions of *SNRPN* exon 1 in early human and mouse embryos result in a paternal to maternal imprint switch. *Nature Genetics* **25**, 74-78 (2000).

25. Paulsen, M., Davies, K.R., Bowden, L.M., Villar, A.J., Franck, O., Fuermann, M., Dean, W.L., Moore, T.F., Rodrigues, N., Davies, K.E., Hu, R.-J., Feinberg, A.P., Maher, E.R., Reik, W. & Walter, J. Syntenic organisation of the mouse distal chromosome 7 imprinting cluster and the Beckwith-Wiedemann syndrome region in chromosome 11p15.5. *Human Molecular Genetics* **7**(1998).

26. Lee, M.P., Brandenburg, S., Landes, G.M., Adams, M., Miller, G. & Feinberg, A.P. Two novel genes in the centre of the 11p15 imprinted domain escape genomic imprinting. *Human Molecular Genetics* **8**, 683-690 (1999).

27. Leighton, P.A., Ingram, R.S., Eggenschwiler, J., Efstratiadis, A. & Tilghman, S.M. Disruption of imprinting caused by deletion of the H19 gene region in mice. *Nature* **375**, 34-39 (1995).

28. John, R.M., Hodges, M., Little, P., Barton, S.C. & Surani, M.A. A human p57KIP2 transgene is not activated by passage through the maternal mouse germline. *Human Molecular Genetics* **8**, 2211-2219 (1999).

29. Brockdorff, N. Convergent themes in X chromosome inactivation and autosomal imprinting. in *Genomic Imprinting* (eds. Reik, W. & Surani, A.) (Oxford University Press, Oxford, New York, Tokyo, 1997).

30. Penny, G.D., Kay, G., F., Sheardown, S.A., Rastan, S. & Brockdorff, N. Requirement for *Xist* in X chromosome inactivation. *Nature* **379**, 131-137 (1996).

31. Clemson, C.M., McNeil, J.A., Willard, H.F. & Lawrence, J.B. XIST RNA Paints the Inactive X Chromosome at Interphase: Evidence for a Novel RNA Involved in Nuclear/Chromosome Structure. *The Journal of Cell Biology* **132**, 259-275 (1996).

32. Lyon, M.F. X-Chromosome inactivation: a repeat hypothesis. *Cytogenetics and Cell Genetics* **80**, 133-137 (1998).

33. Bailey, J.A., Carrel, L., Chakravarti, A. & Eichler, E.E. Molecular evidence for a relationship between LINE-1 elements and X-chromosome inactivation: The Lyon repeat hypothesis. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 6634-6639 (2000).

34. Tremblay, K.D., Saam, J.R., Ingram, R.S., Tilghman, S.M. & Bartolomei, M.S. A paternal-specific methylation imprint marks the alleles of the mouse *H19* gene. *Nature Genetics* **9**, 407-413 (1995).

35. Hatada, I. & Mukai, T. Genomic imprinting of p57KIP2, a cyclin-dependent kinase inhibitor, in mouse. *Nature Genetics* **11**, 204-206 (1995).

36. Ainscough, J.F.X., Koide, T., Tada, M., Barton, S. & Surani, M.A. Imprinting of Igf2 and H19 from a 130 kb YAC transgene. *Development* **124**, 3621-3632 (1997).

37. Elson, D.A. & Bartolomei, M.S. A 5' Differentially Methylated Sequence and the 3' Flanking Region are Necessary for *H19* Transgene Imprinting. *Molecular and Cellular Biology* **17**, 309-317 (1997).

38. Bartolomei, M.S., Webber, A.L., Brunkow, M.E. & Tilghman, S.M. Epigenetic mechanisms underlying the imprinting of the mouse *H19* gene. *Genes and Development* **7**, 1663-1673 (1993).

39. Pfeifer, K., Leighton, P.A. & Tilghman, S.M. The structural *H19* gene is required for transgene imprinting. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 13877-13883 (1996).

40. Pirrotta, V. Chromatin complexes regulating gene expression in *Drosophila*. *Current Opinion in Genetics and Development* **5**, 466-472 (1995).

41. Kitsberg, D., Selig, S., Brandeis, M., Simon, I., Keshet, I., Driscoll, D.J., Nicholls, R.D. & Cedar, H. Allele-specific replication timing of imprinted gene regions. *Nature* **364**, 459-463 (1993).

42. Gunaratne, P.H., Nakao, M., Ledbetter, D.H., Sutcliffe, J.S. & Chinault, A.C. Tissue-specific and allele-specific replication timing control in the imprinted human Prader-Willi syndrome region. *Genes and Development* **9**, 808-820 (1995).

43. Simon, I., Tenzen, T., Reubinoff, B.E., Hillman, D., McCarrey, J.R. & Cedar, H. Asynchronous replication of imprinted genes is established in the gametes and maintained during development. *Nature* **401**, 929-932 (1999).

44. Jamieson, R.V., Tam, P.P.L. & Gardiner-Garden, M. X-chromosome activity: impact of imprinting and chromatin structure. *International Journal of Developmental Biology* **40**, 1065-1080 (1996).

45. Greally, J.M., Starr, D.J., Hwang, S., Song, L., Jaarola, M. & Zemel, S. The mouse *H19* locus mediates a transition between imprinted and non-imprinted DNA replication patterns. *Human Molecular Genetics* **7**, 91-95 (1998).

46. Ariel, M., Selig, S., Brandeis, M., Kitsberg, D., Kafri, T., Weiss, A., Keshet, I., Razin, A. & Cedar, H. Allele-specific Structures in the Mouse *Igf2-H19* Domain. *Cold Spring Harbour Symposium on Quantitative Biology* **LVIII**, 307-313 (1993).

47. Bickmore, W.A. & Carothers, A.D. Factors affecting the timing and imprinting of replication on a mammalian chromosome. *Journal of Cell Science* **108**, 2801-2809 (1995).

48. Yu, Y., Xu, F., Peng, H., Fang, X., Zhao, S., Li, Y., Cuevas, B., Kuo, W.L., Gray, J.W., Siciliano, M., Mills, G.B. & Bast, R.B. *NOEY2* (ARHI), an imprinted putative tumor suppressor gene in ovarian and breast carcinomas. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 214-219 (1999).

49. Kamiya, M., Judson, H., Okazaki, Y., Kusakabe, M., Muramatsu, M., Takada, S., Takagi, N., Arima, T., Wake, N., Kamimura, K., Satomura, K., Hermann, R., Bonthron, D.T. & Hiyashizaki, Y. The cell cycle control gene *ZAC/PLAGL1* is imprinted- a stong candidate for transient neonatal diabetes. *Human Molecular Genetics* **9**, 453-460 (2000).

50. DeChiara, T.M., Efstratiadis, A. & Robertson, E.J. A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting. *Nature* **345**, 78-80 (1990).

51. Barlow, D.P., Stoger, R., Herrman, B.G., Saltow, K. & Schweifer, N. The mouse insulin-like growth factor type-2 receptor is closely linked to the *Tme* locus. *Nature* **349**, 85-87 (1991).

52. Guillemot, F., Nagy, A., Auerbach, A., Rossant, J. & Joyner, A.L. Essential role of *Mash-2* in extraembryonic development. *Nature* **371**, 333-336 (1994).

53. Kobayashi, S., Kohda, T., Miyoshi, N., Kuroiwa, Y., Aisaka, K., Tsutsumi, O., Kaneko-Ishino, T. & Ishino, F. Human *PEG1/MEST*, an imprinted gene on chromosome 7. *Human Molecular Genetics* **6**, 781-786 (1997).

54. Lefebvre, L., Viville, S., Barton, S.C., Ishino, F., Keverne, E.B. & Surani, M.A.H. Abnormal maternal behaviour and growth retardation associated with loss of the imprinted gene *Mest*. *Nature Genetics* **20**, 163-169 (1998).

55. Mayer, W., Hemberger, M., Frank, H.-G., Grummer, R., Winterhager, E., Kaufmann, P. & Fundele, R. Expression of the Imprinted Genes *MEST/Mest* in Human and Murine Placenta Suggests a Role in Angiogenesis. *Developmental Dynamics* **217**, 1-10 (2000).

56. Li, Y. & Behringer, R.R. *Esx1* is an X-chromosome-imprinted regulator of placental development and fetal growth. *Nature Genetics* **20**, 309-311 (1998).

57. Hansen, H., Svensson, U., Zhu, J., Laviola, L., Giorgino, F., Wolf, G., Smith, R.J. & Riedel, H. Interaction between the Grb10 SH2 Domain and the Insulin Receptor Carboxyl Terminus. *The Journal of Biological Chemistry* **271**, 8882-8886 (1996).

58. Duvillie, B., Cordonnier, N., Deltour, L., Dandoy-Dron, F., Itier, J.-M., Monthioux, E., Jami, J., Joshi, R.L. & Bucchini, D. Phenotypic alterations in insulin-deficient mutant mice. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 5137-5140 (1997).

59. Spencer, H.G., Clark, A.G. & Feldman, M.W. Genetic conflicts and the evolutionary origin of genomic imprinting. *Trends in Ecology and Evolution* **14**, 197-201 (1999).

60. Hurst, L.D. Evolutionary theories of genomic imprinting. in *Genomic Imprinting* (eds. Reik, W. & Surani, A.) (Oxford University Press, Oxford, New York, Tokyo, 1997).

61. Moore, T. & Haig, D. Genomic imprinting in mammalian development: a parental tug of war. *Trends in Genetics* **7**, 45-49 (1991).

62. Hurst, L.D. & McVean, G.T. Growth effects of uniparental disomies and the conflict theory of genomic imprinting. *Trends in Genetics* **13**, 436-443 (1997).

63. Kuroiwa, Y., Kaneko-Ishino, T., Kagitani, F., Kohda, T., Li, L.-L., Tada, M. & Surani, M.A. Peg3, an imprinted gene on proximal chromosome 7 encodes for a zinc finger protein. *Nature Genetics* **12**, 186-190 (1996).

64. Kurz, H., Zechner, U., Orth, A. & Fundele, R. Lack of correlation between placental and offspring size in mouse interspecific crosses. *Anatomical Embryology* **200**, 335-343 (1999).

65. Golic, K.G., Golic, M.M. & Pimpinelli, S. Imprinted control of gene activity in *Drosophila*. *Current Biology* **8**, 1273-1276 (1998).

66. Lloyd, V.K., Sinclair, D.A. & Grigliatti, T.A. Genomic Imprinting and Position-Effect Variegation in *Drosophila melanogaster*. *Genetics* **151**, 1503-1516 (1999).

67. Martin, C.C. & McGowan, R. Parent-of-Origin Specific Effects on the Methylation of a Transgene in the Zebrafish, *Danio rerio*. *Developmental Genetics* **17**, 233-239 (1995).

68. McVean, G.T. & Hurst, L.D. Molecular evolution of imprinted genes: no evidence for antagonistic coevolution. *Proceedings of the Royal Society of London, Series B* **264**, 739-746 (1997).

69. Toder, R., Wilcox, S.A., Smithwick, M. & Graves, G.A.M. The human/mouse imprinted genes *IGF2, H19, SNRPN* and *ZNF127* map to two conserved autosomal clusters in a marsupial. *Chromosome Research* **4**, 295-300 (1996).

70. DeChiara, T.M., Robertson, E.J. & Efstratiadis, A. Parental imprinting of the mouse Insulin-like Growth Factor-II gene. *Cell* **64**, 849-859 (1991).

71. Sun, H.-L., Dean, W.L., Kelsey, G., Allen, N.D. & Reik, W. Transactivation of *Igf2* in a mouse model of Beckwith-Wiedemann syndrome. *Nature* **389**, 809-815 (1997).

72. Zhang, P., Liegeois, N.J., Wong, C., Finegold, M., Hou, H., Thompson, J.C., Silverman, A., Harper, J.W., DePinho, R.A. & Elledge, S.J. Altered cell differentiation and proliferation in mice lacking p57KIP2 indicates a role in Beckwith-Wiedemann syndrome. *Nature* **387**, 151-158 (1997).

73. Caspary, T., Cleary, M.A., Perlmann, E.J., Zhang, P., Elledge, S.J. & Tilghman, S.M. Oppositely imprinted genes *p57Kip2* and *Igf2* interact in a mouse model for Beckwith-Wiedemann syndrome. *Genes and Development* **13**, 3115-3124 (1999).

74. Feinberg, A.P. Genomic imprinting as a developmental process disturbed in cancer. in *Genomic Imprinting* (eds. Reik, W. & Surani, A.) (Oxford University Press, Oxford, New York, Tokyo, 1997).

75. Feil, R., Moore, T.F., Oswald, J., Walter, J., Sun, F. & Reik, W. The imprinted insulin-like growth factor 2 gene. in *Genomic Imprinting* (eds. Reik, W. & Surani, A.) (Oxford University Press, Oxford, New York, Tokyo, 1997).

76. Ekstrom, T.J., Cui, H., Li, X. & Ohlsson, R. Promoter-specific *IGF2* imprinting status and its plasticity during human liver development. *Development* **121**, 309-316 (1995).

77. Brannan, C.I., Dees, E.C., Ingram, R.S. & Tilghman, S.M. The Product of the *H19* Gene May Function as an RNA. *Molecular and Cellular Biology* **10**, 28-36 (1990).

78. Pachnis, V., Brannan, C.I. & Tilghman, S.M. The structure and expression of a novel gene activated in early mouse embryogenesis. *The EMBO Journal* **7**, 673-681 (1988).

79. Lin, W.-L., He, X.-B., Svensson, K., Adam, G., Li, Y.-M., Tang, T.-W., Paldi, A., Pfeifer, S. & Ohlsson, R. The genotype and epigenotype synergise to diversify the spatial pattern of expression of the imprinted *H19* gene. *Mechanisms of Development* **82**, 195-197 (1999).

80. Pachnis, V., Belayew, A. & Tilghman, S.M. Locus unlinked to a-fetoprotein under the control of the murine raf and Rif genes. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 5523-5527 (1984).

81. Bartolomei, M.S., Zemel, S. & Tilghman, S.M. Parental imprinting of the mouse *H19* gene. *Nature* **351**, 153-155 (1991).

82. Hurst, L.D. & Smith, N.G.C. Molecular evolutionary evidence that H19 mRNA is functional. *Trends in Genetics* **15**, 134-135 (1999).

83. Li, Y.-M., Franklin, G., Cui, H.-M., Svensson, K., He, X.-B., Adam, G., Ohlsson, R. & Pfeifer, S. The *H19* Transcript Is Associated with Polysomes and May Regulate *IGF2* Expression in *trans*. *The Journal of Biological Chemistry* **43**, 28247-28252 (1998).

84. Ripoche, M.-A., Kress, C., Poirier, F. & Dandolo, L. Deletion of the *H19* transcription unit reveals the existence of a putative imprinting control element. *Genes and Development* **11**, 1596-1604 (1997).

85. Schmidt, J.V., Levorse, J.M. & Tilghman, S.M. Enhancer competition between *H19* and *Igf2* does not mediate their imprinting. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 9733-9738 (1999).

86. Szabo, P.E. & Mann, J.R. Allele-specific expression and total expression levels of imprinted genes during early mouse development - implications for imprinting mechanisms. *Genes & Development* **9**, 3097-3108 (1995).

87. Lee, J.E., Pintar, J. & Efstratiadis, A. Pattern of the insulin-like growth factor-II gene-expression during early mouse embryogenesis. *Development* **110**, 151 (1990).

88. Redline, R.W., Chernicky, C.L., Tan, H.-Q., Ilan, J. & Ilan, J. Differential Expression of Insulin-Like Growth Factor II in Specific Regions of the Late (Post Day 9.5) Murine Placenta. *Molecular Reproduction and Development* **36**, 121-129 (1993).

89. Correia-da-Silva, G., Bell, S.C., Pringle, J.H. & Teixeira, N. Expression of mRNA Encoding Insulin-Like Growth Factors I and II by Uterine Tissues and Placenta During Pregnancy in the Rat. *Molecular Reproduction and Development* **53**, 294-305 (1999).

90. Ohlsson, R., Larsson, E., Nilsson, O., Wahlstrom, T. & Sundstrom, P. Blastocyst implantation precedes induction of insulin-like growth factor II gene expression in human trophoblasts. *Development* **106**, 555-559 (1989).

91. Brice, A.L., Cheetham, J.E., Bolton, V.N., Hill, N.C.W. & Schofield, P.N. Temporal changes in the expression of the insulin-like growth factor II gene associated with tissue maturation in the human fetus. *Development* **106**, 543-554 (1989).

92. Ohlsson, R., Holmgren, L., Glaser, A., Szpecht, A. & Pfeifer-Ohlsson, S. Insulin-like growth factor 2 and short range stimulatory loops in control of human placental growth. *The EMBO Journal* **8**, 1993-1999 (1989).

93. Poirier, F., Chan, C.T.J., Timmons, P.M., Robertson, E.J., Evans, M.J. & Rigby, P.W. The murine H19 gene is activated during embryonic stem cell differentiation in vitro and at the time of implantation in the developing embryo. *Development* **113**, 1105-1114 (1991).

94. Lustig, D., Ariel, I., Ilan, J., Lev-Lehman, E., DeGroot, N. & Hochberg, A. Expression of the Imprinted Gene *H19* in the Human Fetus. *Molecular Reproduction and Development* **38**, 239-246 (1994).

95. Stylianpoulou, F., Efstratiadis, A., Herbert, J. & Pintar, J. Pattern of the insulin-like growth factor II gene expression during rat embryogenesis. *Development* **103**, 497-506 (1988).

96. Schofield, P.N. & Tate, V.E. Regulation of human IGFII transcription in fetal and adult tissues. *Development* **101**, 793-803 (1987).

97. Stylianpoulou, F., Herbert, J., Bento-Soares, M. & Efstratiadis, A. Expression of the insulin-like growth factor II gene in the choroid plexus and the leptomeninges of the adult rat central nervous system. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 141-145 (1988).

98. Svensson, K., Walsh, C., Fundele, R. & Ohlsson, R. H19 is imprinted in the choroid plexus and leptomeninges of the mouse foetus. *Mechanisms of Development* **51**, 31-37 (1995).

99. Hemberger, M., Redies, C., Krause, R., Oswald, J., Walter, J. & Fundele, R.H. H19 and Igf2 are expressed and differentially imprinted in neuroectoderm-derived cells in the mouse brain. *Dev. Genes Evol.* **208**, 393-402 (1998).

100. Hetts, S.W., Rosen, K.M., Dikkes, P., Villa-Komaroff, L. & Mozell, R.L. Expression and Imprinting of the Insulin-Like Growth Factor II Gene in Neonatal Mouse Cerebellum. *Journal of Neuroscience Research* **50**, 958-966 (1997).

101. Ohlsson, R., Hedborg, F., Holmgren, L., Walsh, C. & Ekstrom, T.J. Overlapping patterns of *IGF2* and *H19* expression during human development: biallelic *IGF2* expression correlates with a lack of *H19* expression. *Development* **120**, 361-368 (1994).

102. Overall, M., Bakker, M., Spencer, J., Parker, N., Smith, P. & Dziadek, M. Genomic Imprinting in the Rat: Linkage of *Igf2* and *H19* Genes and Opposite Parental Allele-Specific Expression during Embryogenesis. *Genomics* **45**, 416-420 (1997).

103. Yoo-Warren, H., Pachnis, V., Ingram, R.S. & Tilghman, S.M. 2 regulatory domains flank the mouse H19 gene. *Molecular and Cellular Biology* **8**, 4707-4715 (1988).

104. Leighton, P.A., Saam, J.R., Ingram, R.S., Stewart, C.L. & Tilghman, S.M. An enhancer deletion affects both H19 and Igf2 expression. *Genes & Development* **9**, 2079-2089 (1995).

105. Brunkow, M.E. & Tilghman, S.M. Ectopic expression of the H19 gene in mice causes perinatal lethality. *Genes and Development* **5**, 1092-1101 (1991).

106. Hatano, N., Eversole-Cire, P., Ferguson-Smith, A.C., Jones, P.A., Surani, M.A. & Sasaki, H. Enhancer-Dependent, Locus-Wide Regulation of the Imprinted Mouse Insulin-Like Growth Factor II Gene. *Journal of Biochemistry* **123**, 984-991 (1998).

107. Ainscough, J.F.-X., Dandolo, L. & Surani, M.A. Appropriate expression of the mouse *H19* gene utilises three or more distinct enhancer regions spread over more than 130kb. *Mechanisms of Development* **91**, 365-368 (2000).

108. Ward, A., Fisher, R., Richardson, L., Pooler, J.A., Squire, S., Bates, P., Shaposhnikov, R., Hayward, N., Thurston, M. & Grayham, C.F. Genomic regions regulating imprinting and insulin-like growth factor-II promoter 3 activity in transgenics: novel enhancer and silencer elements. *Genes and Function* **1**, 25-36 (1997).

109. Brenton, J.D., Drewell, R.A., Viville, S., Hilton, K.J., Barton, S.C., Ainscough, J.F.-X. & Surani, M.A. A silencer element identified in *Drosophila* is required for

imprinting of *H19* reporter transgenes in mice. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 9242-9247 (1999).

110. Latham, K.E., Doherty, A.S., Scott, C.D. & Shultz, R.M. *Igf2r* and *Igf2* gene expression in androgenetic, gynogenetic, and parthenogenetic preimplantation mouse embryos: absence of regulation by geneomic imprinting. *Genes and Development* **8**, 290-299 (1994).

111. Jouvenot, Y., Poirier, F., Jami, J. & Paldi, A. Biallelic transcription of *Igf2* and *H19* in individual cells suggests a post-transcriptional contribution to genomic imprinting. *Current Biology* **9**, 1199-1202 (1999).

112. Szabo, P. & Mann, J.R. Expression and methylation of imprinted genes during in vitro differentiation of mouse parthenogenetic and androgenetic embryonic stem cells. *Development* **120**, 1651-1660 (1994).

113. Sasaki, H., Ferguson-Smith, A.C., Shum, A.S.W., Barton, S. & Surani, M.A. Temporal and spatial regulation of *H19* imprinting in normal and uniparental mouse embryos. *Development* **121**, 4195-4202 (1995).

114. Adam, G.A.R., Cui, H., Miller, S.J., Flam, F. & Ohlsson, R. Allele-specific in situ hybridisation (ASISH) analysis: a novel technique which resolves differential allelic usage of *H19* within the same cell lineage during human placental development. *Development* **122**, 839-847 (1996).

115. Ohlsson, R., Flam, F., Fisher, R., Miller, S., Cui, H., Pfeifer, S. & Adam, G.A.R. Random monoallelic expression of the imprinted *IGF2* and *H19* genes in the absence of discriminative parental marks. (1999).

116. Svensson, K., Mattsson, R., James, T.C., Wentzel, P., Pilartz, M., MacLaughin, J., Miller, S.J., Olsson, T., Eriksson, U.J. & Ohlsson, R. The paternal allele of the *H19* gene is progressively silenced during early mouse development: the acetylation of the histones may be involved in the generation of variegated expression patterns. *Development* **125**, 61-69 (1998).

117. Reik, W. & Surani, A. *Genomic Imprinting* (eds. Reik, W. & Surani, A.) (Oxford University Press, Oxford, New York, Tokyo, 1997).

118. Li, E., Bestor, T.H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915-926 (1992).

119. Li, E., Beard, C. & Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* **366**, 362-365 (1993).

120. Tucker, K.L., Beard, C., Dansman, J., Jackson-Grusby, L., Laird, P.W., Lei, H., Li, E. & Jaenisch, R. Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not non-imprinted genes. *Genes and Development* **10**, 1008-1020 (1996).

121. Monk, M., Boubelik, M. & Lehnert, S. Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* **99**, 371-382 (1987).

122. Monk, M. Changes in DNA methylation during mouse embryonic development in relation to X-chromosome activity and imprinting. *Phil. Trans. R. Soc. Lond. B* **326**, 299-312 (1990).

123. Mayer, W., Niveleau, A., Walter, J., Fundele, R. & Haaf, T. Demethylation of the zygotic paternal genome. *Nature* **403**, 501-502 (2000).

124. Bhattacharya, S.K., Ramchandani, S., Cervoni, N. & Szyf, M. A mammalian protein with specific demethylase activity for mCpG DNA. *Nature* **397**, 579-583 (1999).

125. Bird, A.P. CpG-rich islands and the function of DNA methylation. *Nature* **321**, 209-213 (1986).

126. Eden, S. & Cedar, H. Role of DNA methylation in the regulation of transcription. *Current Opinion in Genetics and Development* **4**, 255-259 (1994).

127. Razin, A. CpG methylation, chromatin structure and gene silencing-a three way connection. *The EMBO Journal* **17**, 4905-4908 (1998).

128. Brandeis, M., Kafri, T., Ariel, M., Chaillet, J.R., McCarrey, J., Razin, A. & Cedar, H. The ontogeny of allele-specific methylation associated with imprinted genes in the mouse. *The EMBO Journal* **12**, 3669-3677 (1993).

129. Otte, K., Choudhury, D., Charalambous, M., Engstrom, W. & Rozell, B. A conserved structural element in horse and mouse IGF2 genes binds a methylation sensitive factor. *Nucleic Acids Research* **26**, 1605-1612 (1998).

130. Tremblay, K.D., Duran, K.L. & Bartolomei, M.S. A 5' 2-Kilobase-Pair Region of the Imprinted Mouse *H19* Gene Exhibits Exclusive Paternal Methylation throughout Development. *Molecular and Cellular Biology* **17**, 4322-4329 (1997).

131. Warnecke, P.M., Mann, J.R., Frommer, M. & Clark, S.J. Bisulphite Sequencing in Preimplantation Embryos: DNA Methylation Profile of the Upstream Region of the Mouse Imprinted *H19* Gene. *Genomics* **51**, 182-190 (1998).

132. Tada, T., Tada, M., Hilton, K., Barton, S.C., Sado, T., Takagi, N. & Surani, M.A. Epigenotype switching of imprintable loci in embryonic germ cells. *Development, Genes and Evolution* **207**, 551-561 (1998).

133. Kato, K., Rideout III, W.M., Hilton, K., Barton, S., Tsunoda, Y. & Surani, M.A. Developmental potential of mouse primordial germ cells. *Development* **126**, 1823-1832 (1999).

134. Davis, T.L., Yang, G.J., McCarrey, J. & Bartolomei, M.S. Methylation is acquired at different times on the maternal and paternal *H19* alleles during male germ cell development. *Meeting report, 13th International Mammalian Genome Conference* (1999).

135. Davis, T.L., Trasler, J.M., Moss, S.B., Yang, G.J. & Bartolomei, M.S. Acquisition of the *H19* Methylation Imprint Occurs Differentially on the Parental Alleles during Spermatogenesis. *Genomics* **58**, 18-28 (1999).

136. Villar, A.J., Eddy, E.M. & Pederson, R.A. Developmental Regulation of Genomic Imprinting during Gametogenesis. *Developmental Biology* **172**, 264-271 (1995).

137. Steger, D.J. & Workman, J.L. Remodelling chromatin structures for transcription: what happens to the histones? *BioEssays* **18**, 875-883 (1996).

138. Travers, A. Chromatin modification by DNA tracking. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 13634-13637 (1999).

139. O'Neill, L.P. & Turner, B.M. Histone H4 acetylation distinguishes coding regions of the human genome from heterochromatin in a differentiation-dependent but transcription-independant manner. *The EMBO Journal* **14**, 3946-3957 (1995).

140. Franklin, G.C., Adam, G.I.R. & Ohlsson, R. Genomic Imprinting and Mammalian Development. *Placenta* **17**, 3-14 (1996).

141. Adenot, P.G., Mercier, Y., Renard, J.-P. & Thompson, E.M. Differential H4 acetylation of paternal and maternal chromatin precedes DNA replication and differential transcriptional activity in pronuclei of 1-cell mouse embryos. *Development* **124**, 4615-4625 (1997).

142. Jepperson, P. & Turner, B.M. The Inactive X Chromosome in Female Mammals Is Distinguished by a Lack of Histone H4 Acetylation, a Cytogenetic Marker for Gene Expression. *Cell* **74**, 281-289 (1993).

143. Wade, P.A., Pruss, D. & Wolffe, A.P. Histone acetylation: chromatin in action. *Trends in Biological Sciences* **22**, 128-132 (1997).

144. Struhl, K. Histone acetylation and transcriptional regulatory mechanisms. *Genes and Development* **12**, 599-606 (1998).

145. Nan, X., Ng, H.-H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N. & Bird, A. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386-389 (1998).

146. Jones, P.L., Veenstra, G.J.C., Wade, P.A., Vermaak, D., Kass, S.V., Landsberger, N., Strouboulis, J. & Wolffe, A.P. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature Genetics* **19**, 187-191 (1998).

147. Bird, A.P. & Wolffe, A.P. Methylation-Induced Repression-Belts, Braces, and Chromatin. *Cell* **99**, 451-454 (1999).

148. Nan, X., Tate, P., Li, E. & Bird, A. DNA Methylation Specifies Chromosomal Localisation of MeCP2. *Molecular and Cellular Biology* **16**, 414-421 (1996).

149. Tartof, K.D., Hobbs, C. & Jones, M. A Structural Basis for Variegating Position Effects. *Cell* **37**, 869-878 (1984).

150. Busturia, A., Wightman, C.D. & Sakonju, S. A silencer is required for maintenance of transcriptional repression throughout *Drosophila* development. *Development* **124**, 4343-4350 (1997).

151. Koide, T., Ainscough, J., Wijgerde, M. & Surani, M.A. Comparative-analysis of Igf-2 H19 imprinted domain - identification of a highly conserved intergenic DNase-I hypersensitive region. *Genomics* **24**, 1-8 (1994).

152. Feil, R., Handel, M.A., Allen, N.D. & Reik, W. Chromatin structure and imprinting - developmental control of DNase-I sensitivity in the mouse Insulin-like-Growth-Factor-2 gene. *Developmental Genetics* **17**, 240-252 (1995).

153. Hark, A. & Tilghman, S.M. Chromatin conformation of the *H19* epigenetic mark. *Human Molecular Genetics* **7**, 1979-1985 (1998).

154. Khosla, S., Aitchison, A., Gregory, R., Allen, N.D. & Feil, R. Parental Allele-Specific Chromatin Configuration in a Boundary-Imprinting-Control Element Upstream of the Mouse *H19* Gene. *Molecular and Cellular Biology* **19**, 2556-2566 (1999).

155. Kanduri, C., Holmgren, C., Pilartz, M., Franklin, G., Kanduri, M., Lui, L., Ginjala, V., Ulleras, E., Mattsson, R. & Ohlsson, R. The 5' flank of mouse *H19* in an unusual chromatin conformation unidirectionally blocks enhancer-promoter communication. *Current Biology* **10**, 449-457 (2000).

156. Stadnick, M.P., Pieracci, F.M., Cranston, M.J., Taksel, E., Thorvaldsen, J.L. & Bartolomei, M.S. Role of a 461-bp G-rich element in *H19* transgene imprinting. *Dev. Genes Evol* **209**, 239-248 (1999).

157. Szabo, P.E., Pfeifer, G.P. & Mann, J.R. Characterisation of Novel Parent-Specific Epigenetic Modifications Upstream of the Imprinted Mouse *H19* Gene. *Molecular and Cellular Biology* **18**, 6767-6776 (1998).

158. Lyko, F., Brenton, J.D., Surani, M.A. & Paro, R. An imprinting element from the mouse H19 locus functions as a silencer in Drosophila. *Nature Genetics* **16**, 171-173 (1997).

159. Pedone, P.V., Pikhaart, M.J., Cerrato, F., Vernucci, M., Ungaro, P., Bruni, C.B. & Riccio, A. Role of histone acetylation and DNA methylation in the maintenance of the imprinted expression of the *H19* and *Igf2* genes. *FEBS Letters* **458**, 45-50 (1999).

160. Jones, B.K., Levorse, J.M. & Tilghman, S.M. Igf2 imprinting does not require its own DNA methylation or H19 RNA. *Genes and Development* **12**, 2200-2207 (1998).

161. Thorvaldsen, J.L., Duran, K.L. & Bartolomei, M.S. Deletion of the *H19* differentially methylated domain results in loss of imprinted expression of *H19* and *Igf2*. *Genes and Development* **12**, 3693-3702 (1998).

162. Srivastava, M., Hsieh, S., Grinberg, A., Williams-Simons, L., Huang, S.-P. & Pfeifer, K. *H19* and *Igf2* monoallelic expression is regulated in two distinct ways by a shared *cis* acting regulatory region upstream of *H19*. *Genes and Development* **14**, 1186-1195 (2000).

163. Webber, A.L., Ingram, R.S., Levorse, J.M. & Tilghman, S.M. Location of enhancers is essential for the imprinting of H19 and Igf2 genes. *Nature* **391**, 711-715 (1998).

164. Drewell, R.A., Brenton, J.D., Ainscough, J.F.-X., Barton, S.C., Hilton, K.J., Arney, K.L., Dandolo, L. & Surani, M.A. Deletion of a silencer element disrupts H19 imprinting independantly of a DNA methylation epigenetic switch. *Development* **127**, 3419-3428 (2000).

165. Sun, F.-L. & Elgin, S.C.R. Putting Boundaries on Silence. *Cell* **99**, 459-462 (1999).

166. Kim, J., Shen, B., Rosen, C. & Dorsett, D. The DNA-Binding and Enhancer-Blocking Domains of the *Drosophila suppressor of Hairy-wing* Protein. *Molecular and Cellular Biology* **16**, 3381-3392 (1996).

167. Bi, X. & Broach, J.R. UASrpg can function as a heterochromatin boundary element in yeast. *Genes and Development* **13**, 1089-1101 (1999).

168. Chung, J.H., Bell, A.C. & Felsenfeld, G. Characterisation of the chicken β-globin insulator. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 575-580 (1997).

169. Hark, A., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M. & Tilghman, S.M. CTCF mediates methylation-sensitive enhancer-blocking activity at the *H19/Igf2* locus. *Nature* **405**, 486-489 (2000).

170. Bell, A.C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* **405**, 483-485 (2000).

171. Bell, A.C., West, A.G. & Felsenfeld, G. The Protein CTCF Is Required for the Enhancer Blocking Activity of Vertebrate Insulators. *Cell* **98**, 387-396 (1999).

172. Cai, H.N. & Levine, M. The gypsy insulator can function as a promoter-specific silencer in the *Drosophila* embryo. *The EMBO Journal* **16**, 1732-1741 (1997).

173. Lee, J.E., Tantravahi, U., Boyle, A.L. & Efstratiadis, A. Parental Imprinting of an *Igf-2* Transgene. *Molecular Reproduction and Development* **35**, 382-390 (1993).

174. Hu, J.-F., Vu, T.H. & Hoffman, A.R. Genomic Deletion of an Imprint Maintenance Element Abolishes Imprinting of Both Insulin-Like Growth Factor II and *H19*. *The Journal of Biological Chemistry* **272**, 20715-20720 (1997).

175. Dell, G., Ward, A. & Engstrom, W. Regulation of a promoter from the mouse insulin like growth factor II gene by glucocorticoids. *FEBS Letters* **419**, 161-165 (1997).

176. Constancia, M., Dean, W., Lopes, S., Moore, T., Kelsey, G. & Reik, W. Deletion of a silencer element in the *Igf2* gene results in loss of imprinting independent of *H19*. (unpublished).

177. Okutsu, T., Kuroiwa, Y., Kagitani, F., Kai, M., Aisaka, K., Tsutsumi, O., Kaneko, Y., Yokomori, K., Surani, M.A., Kohda, T., Kaneko-Ishino, T. & Ishino, F. Expression and Imprinting Status of Human *PEG8/IGF2AS*, a Paternally Expressed Antisense Transcript from the *IGF2* Locus in Wilms' Tumors. *Journal of Biochemistry* **127**, 475-483 (2000).

178. Caricasole, A. & Ward, A. Transactivation of mouse Insulin-like Growth Factor-II (Igf-II) gene promoters by the AP-1 complex. *Nucleic Acids Research* **21**, 1873-1879 (1993).

179. Hogan, B., Beddington, R. & Constantini, F. *Manipulating the Mouse Embryo: a laboratory manual (2nd Edition)*, (Cold Spring Harbour Laboratory Press, 1994).

180. Clarke, A.R., Purdie, C.A., Harrison, D.J., Morris, R.G., Bird, C.C., Hooper, M.L. & Wyllie, A.H. Thymocyte apoptosis induced by p53-dependent and independent pathways. *Nature* **362**, 849-852 (1993).

181. Harlow, E. & Lane, D. *Antibodies, A Laboratory Manual*, (Cold Spring Harbour Laboratory Press, 1988).

182. Cuevas, E.C., Bateman, A.C., Wilkins, B.S., Johnson, P.A., Williams, J.H., Lee, A.H.S., Jones, D.B. & Wright, D.H. Microwave antigen retrieval in immunocytochemistry: a study of 80 antibodies. *The Journal of Clinical Pathology* **47**, 448-452 (1994).

183. Bradford, M.M. A rapid and sensitive method for the quantitation of nanogram quantities of protein utilising the principal of protein dye binding. *Analytical Biochemistry* **72**, 248-254 (1976).

184. Campbell, R.C. *Statistics for Biologists, Second Edition*, (Cambridge University Press, 1967).

185. Sambrook, J., Fritsch, E.F. & Maniatis, T. *Molecular Cloning- A Laboratory Manual*, (Cold Spring Harbour Laboratory Press, 1989).

186. Church, G.M. & Gilbert, W. Genomic sequencing. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **81**, 1991-1995 (1984).

187. Yaffe, D. & Saxel, O. Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. *Nature* **270**, 725-727 (1977).

188. Silberstein, L., Webster, S.G., Travis, M. & Blau, H.M. Developmental Progression of Myosin Gene Expression in Cultured Mouse Cells. *Cell* **46**, 1075-1081 (1986).

189. Norton, P.A. & Coffin, S.M. Bacterial beta-galactosidase as a marker for Rous Sarcoma Virus gene expression and replication. *Molecular and Cellular Biology* **5**, 281-290 (1985).

190. Hoffmann, A., Gath, U., Gross, G., Lauber, J., Getzlaff, R., Hellwig, S., Galla, H.-J. & Conradt, H.S. Constitutive Secretion of b-Trace Protein by Cultivated Porcine Choroid Plexus Epithelial Cells: Elucidation of Its Complete Amino Acid and cDNA Sequences. *Journal of Cellular Physiology* **169**, 235-241 (1996).

191. Freshney, R.I. *Culture of Animal Cells: A Manual of Basic Technique (Second Edition)*, (Alan R. Liss, Inc., New York, 1987).

192. Crook, R.B., Kasagami, H. & Prusiner, S.B. Culture and Characterisation of Epithelial Cells from Bovine Choroid Plexus. *Journal of Neurochemistry* **37**, 845-854 (1981).

193. Isaacs, H.V., Tannahill, D. & Slack, J.M.W. Expression of a novel FGF in the *Xenopus* embryo- a new candidate inducing factor for mesoderm formation and anteroposterior specification. *Development* **114**, 711-720 (1992).

194. Rathjen, P.D., Nichols, J., Toth, S., Edwards, D.R., Heath, J.K. & Smith, A.G. Developmentally programmed induction of differentiation inhibiting activity and the control of stem cell populations. *Genes and Development* **4**, 2308-18 (1990).

195. Yan, C., Costa, R.H., Darnell Jr, J.E., Chen, J. & Van Dyke, T.A. Distinct positive and negative elements control the limited hepatocyte and choroid plexus expression of transthyretin in transgenic mice. *The EMBO Journal* **9**, 869-878 (1990).

196. Jeffreys, A.J., Neumann, R. & Wilson, V. Repeat unit sequence variation in minisattelites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. *Cell* **60**, 473-485 (1990).

197. Malcomson, R.D.G., Clarke, A.R., Peter, A., Coutts, S.B., Howie, S.E.M. & Harrison, D.J. Apoptosis induced by g-irradiation, but not CD4 ligation, of peripheral T lymphocytes *in vivo* is p53 dependent. *Journal of Pathology* **181**, 166-171 (1997).

198. Harats, D., Kurihara, H., Belloni, P., Oakley, H., Ziober, A., Ackley, D., Cain, G., Kurihara, Y., Lawn, R. & Sigal, E. Targeting Gene Expression to the Vascular Wall in Transgenic Mice Using the Murine Preproendothelin-1 Promoter. *The Journal of Clinical Investigation* **95**, 1335-1344 (1995).

199. Shubeita, H.E., Thorburn, J. & Chien, K.R. Microinjection of Antibodies and Expression Vectors Into Living Myocardial Cells. *Circulation* **85**, 2236-2246 (1992).

200. Lee, K.J., Ross, R.S., Rockman, H.A., Harris, A.N., O'Brien, T.X., vanBilsen, M., Shubeita, H.E., Kandolf, R., Brem, G., Price, T., Evans, S.M., Zhu, H., Franz, W.M. & Chien, K.R. Myosin Light Chain 2 Luciferase transgenic mice reveal distinct regulatory programs for cardiac and skeletal specific expression of a single contractile protein gene. *Journal of Biological Chemistry* **267**, 15875-15885 (1992).

201. Alam, J. & Cook, J.L. Reporter Genes: Application to the Study of Mammalian Gene Transcription. *Analytical Biochemistry* **188**, 245-254 (1990).

202. Rugh, R. *The Mouse*, 430 (Oxford University Press, Oxford, 1990).

203. Davson, H. *Physiology of the Cerebrospinal Fluid*, (J. & A. Churchill Ltd., London, 1967).

204. Soares, M.B., Turken, A., Ishii, D., Mills, L., Episkopou, V., Cotter, S., Zeitlin, S. & Efstratiadis, A. The Rat Insulin-like Growth Factor II Gene. *The Journal of Molecular Biology* **192**, 737-752 (1986).

205. Milot, E., Strouboulis, J., Trimborn, T., Wijgerde, M., deBoer, E., Langevelde, A., TanUn, K., Vergeer, W., Yannoutsos, N., Grosveld, F. & Fraser, P. Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell* **87**, 105-114 (1996).

206. Festenstein, R., Toliani, M., Corbella, P., Mamaliki, Parrington, J., Fox, M., Miliou, A., Jones, M. & Kioussis, D. Locus Control Region function and heterochromatin-induced position effect variegation. *Science* **271**, 1123-1125 (1996).

207. Dobie, K., Mehtali, M., McClenaghan, M. & Lathe, R. Variegated gene expression in mice. *Trends in Genetics* **13**, 127-130 (1997).

208. Mehtali, M., LeMeur, M. & Lathe, R. The methylation-free status of a housekeeping gene is lost at high copy number. *Gene* **91**, 179-184 (1990).

209. Swain, J.L., Stewart, T.A. & Leder, P. Parental Legacy Determines Methylation and Expression of an Autosomal Transgene: A Molecular Mechanism for Parental Imprinting. *Cell* **50**, 719-727 (1987).

210. Allen, N.D., Norris, M.L. & Surani, M.A. Epigenetic Control of Transgene Expression and Imprinting by Genotype-Specific Modifiers. *Cell* **61**, 853-861 (1990).

211. Chaillet, J.R., Bader, D.S. & Leder, P. Regulation of genomic imprinting by gametic and embryonic processes. *Genes. Dev* **9**, 1177-1187 (1995).

212. Tollefson, S.E., Lajara, R., McCusker, R.H., Clemmons, D.R. & Rotwein, P. Insulin-like Growth Factors (IGF) in Muscle Development. *The Journal of Biological Chemistry* **264**, 13810-13817 (1989).

213. Tollefson, S.E., Sadow, J.L. & Rotwein, P. Coordinate expression of insulin-like growth factor II and its receptor during muscle differentiation. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 1543-1547 (1989).

214. Kou, K. & Rotwein, P. Transcriptional activation of the Insulin-Like Growth Factor-II Gene during Myoblast Differentiation. *Molecular Endocrinology* **7**, 291-302 (1993).

215. Thornwall, M., Chhajlani, V., Ve Greves, P. & Nyberg, F. Detection of growth hormone receptor mRNA in an ovine choroid plexus epithelium cell line. *Biochemical and Biophysical Research Communications* **217**, 349-353 (1995).

216. Agnarsdottir, G., Thorsteinsdottir, H., Okarsson, T., Matthiasdottir, S., St. Haflidadottir, B., Andresson, O.S. & Andresdottir, V. The long terminal repeat is a determinant of cell tropism of maedi-visna virus. *Journal of General Virology* **81**, 1901-1905 (2000).

217. Tsutsumi, M., Skinner, M.K. & Sanders-Bush, E. Transferrin Gene Expression and Synthesis by Cultured Choroid Plexus Epithelial Cells. *The Journal of Biological Chemistry* **264**, 9626-9631 (1989).

218. Southwell, B.R., Duan, W., Alcorn, D., Brack, C., Richardson, S.J., Kohrle, J. & Schreiber, G. Thyroxine Transport to the Brain: Role of Protein Synthesis by the Choroid Plexus. *Endocrinology* **133**, 2116-2126 (1993).

219. Ramanathan, V.K., Hui, A.C., Brett, C.M. & Giacomini, K.M. Primary Cell Culture of the Rabbit Choroid Plexus: An Experimental System to Investigate Membrane Transport. *Pharmaceutical Research* **13**, 952-954 (1996).

220. Zheng, W., Zhao, Q. & Graziano, J.H. Primary Culture of Choroidal Epithelial Cells: Characterisation of an *in vitro* model of Blood-CSF Barrier. *In Vitro Cellular and Developmental Biology-Animal* **34**, 40-45 (1998).

221. Thomas, T., Stadler, E. & Dziadek, M. Effects of the Extracellular Matrix on Fetal Choroid Plexus Epithelial Cells: Changes in Morphology and Multicellular Organisation Do Not Affect Gene Expression. *Experimental Cell Research* **203**, 198-213 (1992).

222. Symonds, H., Krall, L., Remington, L., Saenz-Robles, M., Lowe, S., Jacks, T. & Van Dyke, T. p53-Dependent Apoptosis Suppresses Tumor Growth and Progression In Vivo. *Cell* **78**, 703-711 (1994).

223. Thomas, T. & Dziadek, M. Capacity to form choroid plexus-like cells in vitro is restricted to specific regions of the mouse neural ectoderm. *Development* **117**, 253-262 (1993).

224. Dickson, P.W., Howlett, G.J. & Schreiber, G. *Journal of Biological Chemistry* **260**, 8214-8219 (1985).

225. Murakami, T., Yasuda, Y., Mita, S., Maeda, S., Shimada, K., Fujimoto, T. & Araki, S. Prealbumin expression during mouse development studied by in situ hybridisation. *Cell Differentiation* **22**, 1-9 (1987).

226. Thomas, T., Power, B., Hudson, P., Schreiber, G. & Dziadek, M. The Expression of Transthyretin mRNA in the Developing Rat Brain. *Developmental Biology* **125**, 415-427 (1988).

227. Wakasugi, S., Maeda, S. & Shimada, K. Structure and expression of the mouse prealbumin gene. *The Journal of Biochemistry* **100**, 49-58 (1986).

228. Kita, H., Kawamoto, S., Okobo, K. & Matsubara, K. Isolation of a prealbumin cDNA from mouse choroid plexus. *NCBI Accession Number gi:7305598* (unpublished).

229. Ainscough, J.F.-X., John, R.M., Barton, S.M. & Surani, M.A. A skeletal muscle-specific mouse *Igf2* repressor lies 40kb downstream of the gene. *Development* **127**, 3923-3930 (2000).

230. Kuerbitz, S.J., Plunkett, B.S., Walsh, W.V. & Kastan, M.B. Wild-type p53 is a cell cycle checkpoint determinant following irradiation. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 7491-7495 (1992).

231. Saenz Robles, M.T., Symonds, H., Chen, J. & Van Dyke, T. Induction versus progression of brain tumor development : differential functions of the pRB and p53-targetting domains of SV40 T-antigen. *Molecular Cell Biology* **14**, 2686-2698 (1994).

232. Smith, N.G.C. & Hurst, L.D. The Causes of Synonymous Rate Variation in the Rodent Genome: Can Substituion Rates be Used to Estimate the Sex Bias in Mutation Rate? *Genetics* **152**, 661-673 (1999).

233. O'Neill, M.J., Ingram, R.S., Vrana, P.B. & Tilghman, S.M. Allelic expression of *IGF2* in marsupials and birds. *Development, Genes and Evolution* **210**, 18-20 (2000).

234. Ishihara, K., Hatano, N., Furuumi, H., Kato, R., Iwaki, T., Miura, K., Jinno, Y. & Sasaki, H. Comparative genomic sequencing identifies novel tissue-specific enhancers and sequence elements for methylation senitive factors implicated in *Igf2/H19* imprinting. *Genome Research* **10**(2000).

235. Barletta, J.M., Rainer, S. & Feinburg, A.P. Reversal of loss of imprinting in tumor cells by 5-Aza-2'-deoxycytidine. *Cancer Research* **57**, 48-50 (1997).

236. Battle, T., Preisser, L., Marteau, V., Meduri, G., Lambert, M., Nitchke, R., Brown, P.D. & Corman, B. Vasopressin V1a Receptor Signalling in a Rat Choroid Plexus Cell Line. *Biochemical and Biophysical Research Communications* **275**, 322-327 (2000).